Dipartimento di Informatica, Bioingegneria,
Robotica ed Ingegneria dei Sistemi

Multisensory learning in adaptive interactive systems

by

Erica Volta

Theses Series **DIBRIS-TH-2020-XX**

DIBRIS, Università di Genova

Via Opera Pia, 13 16145 Genova, Italy http://www.dibris.unige.it/

Università degli Studi di Genova

Dipartimento di Informatica, Bioingegneria,

Robotica ed Ingegneria dei Sistemi

Ph.D. Thesis in Computer Science and Systems Engineering
Computer Science Curriculum

Multisensory learning in adaptive interactive systems

by

Erica Volta

March, 2020

Dottorato di Ricerca in Informatica ed Ingegneria dei Sistemi
Indirizzo Informatica
Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi
Università degli Studi di Genova

DIBRIS, Univ. di Genova
Via Opera Pia, 13
I-16145 Genova, Italy
`http://www.dibris.unige.it/`

Ph.D. Thesis in Computer Science and Systems Engineering
Computer Science Curriculum
(S.S.D. INF/01)

Submitted by Erica Volta
DIBRIS, Univ. di Genova

. . . .

Date of submission: December 2019

Title: Multisensory learning in adaptive interactive systems

Advisor: Prof. Gualtiero Volpe
Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi
Università di Genova

. . .

Ext. Reviewers:
Dott.ssa Ilaria Cutica
Dipartimento di Oncologia ed Emato-Oncologia
Università degli Studi di Milano

. . .

Dott. Antonio Rodá
Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Padova

Abstract

The main purpose of my work is to investigate multisensory perceptual learning and sensory integration in the design and development of adaptive user interfaces for educational purposes. To this aim, starting from renewed understanding from neuroscience and cognitive science on multisensory perceptual learning and sensory integration, I developed a theoretical computational model for designing multimodal learning technologies that take into account these results. Main theoretical foundations of my research are multisensory perceptual learning theories and the research on sensory processing and integration, embodied cognition theories, computational models of non-verbal and emotion communication in full-body movement, and human-computer interaction models. Finally, a computational model was applied in two case studies, based on two EU ICT-H2020 Projects, "weDRAW" and "TELMI", on which I worked during the PhD.

This work is dedicated to you who first taught me the value of listening, which is inherent in our relationship with music, and how much unexpected power, relational and therapeutic, is hidden in the ineffable sound. Much of what I have become, for better or worse, I owe to you.

Acknowledgements

Besides my advisor, without which none of these last four years of life and work would have been possible, I would like to thank my fellow colleagues of InfoMus Lab for the stimulating discussions, for the sleepless nights we were working together before deadlines or during our travelling together, and for all the fun we have had in the last four years now. My sincere thanks also goes to the colleagues of UCL UCLIC and KnowledgeLab teams, RCM, UPF and IIT UVIP teams for the amazing jointly work done. I learnt from all of you. I would like to thank my family: my parents for supporting me spiritually and economically throughout all my educational career, even when they disagreed or did not understand the meaning of my choices. You have always supported me and this result is as much mine as yours. Last but not least, I would like to thank my partner for supported (and endured) me spiritually and daily during these years of complex joint decisions. Thank you for accepting the chaos of a life in constant becoming, together the colours are definitely more vivid.

Table of Contents

Introduction

Technology is nowadays increasingly spread and used in every aspect of life, included education. Whereas it exists a vast literature on the positive use of multisensory applications and technologies to enhance learning experience, current technologies do not sufficiently ground on psychophysics and neuroscientific evidence of multisensory integration and learning. However, several neuroscientific and cognitive studies have focused on uni- and multisensory learning of stimuli [Ahissar and Hochstein, 2004], [Fahle, 2005], stressed the cognitive mechanic of learning and the benefits of multisensory stimulation. Furthermore, the educational model widespread in our schools is based on the quite exclusive use of the visual channel, leaving often no room for the other sensory modalities. Perception has traditionally been viewed as a modular function where the different sensory modalities operates in an independent way, but recent results from psychophysics and cognitive science [Gori et al., 2008], [Barrett and Newell, 2015] have shown that children have a preferential sensory channel to learn specific concepts, as spatial or temporal tasks. In adult sensory plasticity and perceptual learning are considerably more restricted, but even present [Fahle et al., 2002]. Psychologists and philosophers have long argued for the important role of sensory-motor integration and interaction with the world for the cognitive development of subjects [Piaget, 1972], [Vygotsky, 1978b]. The theories, that have most influenced on one hand multisensory processing and on the other hand multimodal interaction design, are Gestalt Theory, cognitive theories on executive functions as Working Memory [Baddeley, 2003] and Activity Theory [Oviatt et al., 2017]. From a technological point of view, in the last decades multimodal applications are widespread, due to their usability that often represents a good example of cognitive ergonomics, let the user to interact using different modalities and shift among them along the interaction. Nevertheless, in the interaction design, often these recent findings on multisensory integration and processing are not taken into consideration, preventing the development of adaptive applications that can really enhance the user's experience. In the next paragraphs, we review the several elements and disciplines' fields that are relevant to design multisensory applications, particularly for educational purposes in primary school. Then we introduce the theoretical foundation of multisensory technology, by taking into consideration several cognitive and psychology theories that have influenced this computer science area. We then go through the neuroscientific evidences on multisensory processing and integration, and the pedagogical framework for active and embodied learning. The second part of this section is dedicated to exploring the state-of-art of multimodal and kinesthetic technology and game design for learning. Then, finally, in the second section of the dissertation, we present two H2020 ICT European projects on multimodal technology for education, applied on two different contexts: multisensory embodied mathematics education for primary school children, both typically developed and sensory impaired (visual impaired and dyslexic) ones, and music learning technology. Both the projects were based on large research consortium partners, who worked closely together throughout the whole projects' duration. For each task or study presented in this work, we mentioned the collaboration with each specific partner involved and how each partner contribute to the task or research result. Nevertheless, the author intend to specify that, in weDRAW, the contribution was generally aimed at designing and writing research protocols for specific psychophysical tasks; participating in data collection; carrying out interaction design processes during the project, with and without stakeholder participation; analysing non-verbal behaviour of children involved in mathematical tasks' learning in order to develop a model of automatic classification of cognitive states during learning activities; developing some of the activities presented; designing the validation protocols of the technologies developed and to analyse the data. As far as the TELMI project is concerned, the contribution was addressed to the data collection for the creation of the multimodal archive on which the project was based; to carry out the data analyses presented and to develop the interaction design with interested stakeholders and colleagues for the development of the user interface presented here; finally, a questionnaire with semi-structured interview was developed to evaluate the effectiveness of the design developed. Also in this case, as in weDRAW, the partners who collaborated with us in each task were mentioned and their contribution was specified.

Part I

Theoretical foundations

Chapter 1

Multisensory perception and integration: why senses matter

1.1 Multisensory perceptual learning and cross-modal generalization

1.1.1 Multisensory plasticity in brain development

Studies on perceptual learning have often focused on learning of unisensory stimuli, but our experience involves constantly multisensory stimulation. Visual and auditory information are integrated in performing so many tasks that involve, for example, space localization, shapes recognition and objects movement, that their integration is one of the most relevant for our knowledge development. A lot of study on human and animal brains have shown that during the early development if it occurs an alteration of the environment or in one sense processing, the result can be a much higher degree of plasticity between senses at the neural level [Shimojo and Shams, 2001], [Majewska and Sur, 2006], [Sleigh and Lickliter, 1997], [Bavelier and Neville, 2002]. Furthermore, it has been argued that multisensory stimulation can provide a redundancy that is crucial to let the subject extract information that otherwise cannot be if the stimulation would be unisensory alone in early infancy. For example, in the Bahrick and Lickliter study [Bahrick and Lickliter, 2000] it has been showed that only if 5 months-old infants are habituated with auditory-visual rhythm presentations, and they are not just habituated with visual-only or auditory-only presentation of rhythm, they can discriminate visually presented patterns. It is generally known that human brain has the capacity to alter and differ its response to interact with the environment. There are not so many studies that related brain changes to behavioral changes in the developmental process, questioning whether neural differences studied in adults' brain are product of "nature" (saying with that biological brain changes and dispositions) or "nurture" (meaning training conduct in early stage of development). There is an interesting study of 2009 [Hyde et al., 2009] that investigate this aspect, showing that with just 15 months of musical training in early childhood, they found structural brain changes correlating with musical, motor and auditory relevant improvements. Two more suggestions from neuroscience are relevant to understand the central role of multisensory (and its narrow) in the developmental brain, and, therefore, relevant for our purpose of study multisensory and multimodal technology for education. One suggestion is based on the comparative study on brain plasticity between musicians and non-musicians and on sensory enhancement of children involved in music training. Playing an instrument, indeed, requires a lot of skills, including reading comprehension of a complex symbolic system as the music notation is, and translating it in a sequential motor activity based on multimodal feedback; developing such fine motor skills; memorizing long musical passages; improvising. Several studies have explored brain based on these high specialized sensorimotor, auditory, and auditory-spatial skills [Amunts et al., 1997], [Schlaug, 2001], [Gaser and Schlaug, 2003], [Gaab and Schlaug, 2003b]. As stated in different researches, professional keyboard players, have significantly more grey matter in several brain regions, including primary sensorimotor cortex and Heschl's gyrus (primary auditory cortex), the cerebellum, the inferior frontal gyrus, and part of the lateral inferior temporal lobe, than either amateur musicians and non-musicians. It may not be surprising that researchers found structural differences in brain regions closely linked to skills learned during music training (such as independent fine motor movements in both hands and auditory discrimination), but the findings on structural differences outside these primary regions (e.g., inferior frontal gyrus) are interesting since this may indicate that plasticity can occur in brain regions that either have control over primary musical functions or serve as multimodal integration regions for musical skills. There are evidences that suggest that the structural differences in musicians' brains are more pronounced if they have begun to study music at a younger age [Schlaug et al., 1995], [Lee et al., 2003] and who practiced with greater intensity [Gaser and Schlaug, 2003], [Schneider et al., 2002]. Furthermore, there are also animal studies on long-term motor training, that support the argument of the association between training and brain plasticity [Anderson et al., 2002]. In [Schlaug et al., 2005], the authors found cognitive and brain effects in young children (5-7 years old) due to musical training, particularly related to fine motor and melodic discrimination, with a correlation with older children's musical skills, supporting the idea of correlation between music training and brain plasticity in developmental ages. One other suggestion is given by studies on blind brain [Pissaloux and Velazquez, 2017]: in [Harrar et al., 2018], the authors, analysing the reorganization of blind brain, found a great plasticity of it, able to cross the usual sensory boundaries. It was found, indeed, that congenitally blind people use the occipital cortex for tactile, auditory and olfactory stimuli interpretation, according to other studies that also revealed that occipital cortex is not solely reserved for visual processing. These studies support the hypothesis that instead of being sensory-specific, cortical regions might be functionally specific, opening new prospective on current rehabilitation framework. This rapid introduction into neuroscientific findings has the relevance of giving a new point of view also to the scientific research on the use of multisensory processing skill for education and developing of adaptive interactive interfaces for children learning and experience, as we will see in the next paragraphs.

Although the perceptual learning in adults is considerably more restricted compared to the developmental brain stages, recent studies have compared auditory-visual and visual learning, using a coherent motion detection and discrimination tasks [Seitz et al., 2006]. In this study, researchers have shown that there is a substantial advantage of auditory-visual training comparing to the visual alone; this type of training let reduce the number of sessions required for learning, while also raising the maximum performance. Interesting results are found also in a complementary study of auditory recognition [Von Kriegstein and Giraud, 2006]; indeed, it has been found that voice recognition was improved by audiovisual training, co-presenting video and audio. The results were compared with the training with auditory stimulus alone. A training on any pair of multisensory stimuli might induce a more effective representation of the unisensory stimulus, as shown in [Kim et al., 2008]. Kim, Seitz and Shams, in this work, compared learning across three groups, one trained with visual motion stimuli only, one with congruent auditory-visual (both moving in the same direction) and one with inconsistent auditory-visual stimuli (i.e. moving in opposite direction). They found that facilitation was specific to the congruent condition, concluding that facilitation involves multisensory interactions. As highlighted in [Shams and Seitz, 2008], many multisensory congruencies can only exist as a function of our experience. For example, a congruency such as a dog barking is "woof" is experientially driven. At the core of the difference between multisensory and unisensory training is the encoding mechanism, because a larger set of processing structures are activated in the multisensory paradigm.

Several studies show that adults integrate visual and haptic information in a statistically optimal way [Ernst and Banks, 2002]: Ernst and Banks show in their work that when a person looks at an object while exploring it with their hands, vision and touch provide information for estimating the properties of it. The most relevant result is the understating that visual often dominates the integrated visual-haptic percept, for example judging a size, shape or position, but at the same time there are circumstances where the percept is clearly affected by haptics. This psychophysics framework, however, can be used to think the multisensory integration just after the 8-10 years old, because recent studies as [Gori et al., 2008] show that in children younger than 8 years old, one sense dominates the other. Prior to 8 years of age, in fact, integration of visual and haptic spatial information is far from optimal, with either vision or touch dominating totally, even if the predominant sense is far less precise than the other and it depends on the task: for size judgement, for example, the haptic system dominates vision, but for orientation and space knowledge is the vision that dominates the haptic modality.

Moreover, research on memory shows that multisensory exposure can facilitate a best object recognition compared to the unisensory one. It has been suggested, for example, that with auditory-visual exposure the memory capacity can be enhanced [Luriia, 1987]. Similarly, Lehman and Murray [Lehmann and Murray, 2005] proposed a recognition task in which subjects had to indicate for each stimulus presentation whether that stimulus had been previously presented; a subset of these stimuli were paired with sounds that corresponded to the image identity (e.g.

image of a bell and “dong”). Even when the sound was not presented the second time, they found that the images that were previously presented with the corresponding sound were better recognized than those that had been presented just visually. In this case the semantic congruency has an important role in multisensory facilitation, with no memory enhancement in inconsistent stimuli cases. In this and other works, as [Horowitz and Prytulak, 1969], [Nyberg et al., 2000], [Murray et al., 2004, Murray et al., 2005], there are evidences, sustained with neuroimaging studies, that a multisensory representation is possible if there can be a multisensory exposure and, thus, this will activate a larger network of brain areas, that can be invoked after unisensory encoding.

1.1.2 Auditory information processing and multisensory integration

In our everyday experience, we navigate through a lot of different sounds characterized by a complex structure, different frequencies and temporal modulations, and our auditory system needs to process the same sounds in different contexts, preserving the ability to separate the sounds from the acoustic scene. In last years, auditory neuroscience has made substantial progress in understanding how sounds are represented in our brain under different conditions, demonstrating that auditory processing of seemingly acoustic features, like frequency and time, is highly dependent on co-occurring acoustic and behavioural stimuli. In auditory processing, at first stage the cochlea decomposed the input sound waveform into electrical signals for different frequency bands, following a frequency-delimited organization characteristic of the auditory processing centres. Therefore, inferior colliculus, auditory thalamus and auditory cortex have a tonotopic organization, that has been considered a fundamental feature of auditory system. Auditory neuroscience, in fact, used pure frequency tones to systematically characterize the response properties of the auditory system. However, natural acoustic stimuli, like speech or environmental sounds, are comprised of complex signals in multiple frequency bands. Encoding complicated spectral profiles is behaviourally important, as these sounds can provide cues for identifying speakers and different sound localisation. Understanding how a complex sound is encoded is not immediately evident by looking at the responses to each frequency components. Responses to distinct spectral components, in fact, interact with each other in frequency and time. Beyond spectral representation, auditory system has nonlinear sensitivity to temporal and spectral context, employing network-level mechanisms, such as cross-band and temporally adaptive inhibition, to modulate stimulus responses across time and frequency. A relevant aspect to consider is, then, the behavioural one: arousal state and reward have substantial effect on sensory responses, revealing highly plastic stimulus representations [Angeloni and Geffen, 2018]. It is particularly interesting to consider auditory functioning in multisensory context and integration. There is a nourished literature growing in the field of multisensory neuroscience investigating the influence of each sensory stimulus to the others. Numerous reports, for example, suggested that blind individuals are able to develop heightened auditory spatial abilities [Gougoux et al., 2004], [Muchnik et al., 1991], [Muchnik et al., 1991], [Niemeyer and Starlinger, 1981]. Auditory spatial processing was particularly investigated due its high relevance in spatial navigation. In literature can be found numerous articles on enhanced sound localization abilities in early blind individuals [Doucet et al., 2005], [Gougoux et al., 2005] and enhanced auditory spatial discrimination abilities [Voss et al., 2004] in the horizontal plane. However, most of these studies compared blind performance to the one of blindfolded sighted individuals. This procedure can lead to several biases in performance; blind people, in fact, may have had the entire life to get used to their condition, while sighted individuals can be severely disadvantaged trying to localize sound without vision. In [Tabry et al., 2013] the authors compared sound localization ability of sighted individuals with and without be blindfolded in a hemi-anechoic chamber. Blind-folded significantly affected all conditions, but subjects had a worse performance for head-pointing in the horizontal plane. Moreover, blindfolded increased the tendency to undershoot more eccentric spatial positions for head-pointing but not in hand-pointing. These findings indicate that head-pointing relies more heavily on visual cues to provide a precise feedback, than hand-pointing. Vision is essential in the domain of space perception influencing also the other sensory modalities. Auditory space perception is modulated by visual inputs, such as that if auditory and visual stimuli are simultaneously presented in different space locations, the auditory stimulus is localized by the location of the visual one. This is known as Ventriloquist effect [Bertelson and Radeau, 1981]. Unlike the visual system, the auditory system cannot rely on retinotopic organization of space and for such reason is normally less accurate than the visual system. Vision can also interact with audition even when the visual stimulus is not provided during an auditory task [Tabry et al., 2013]. In [Tonelli et al., 2015] the authors provide a demonstration of auditory space bisection in sighted people, demonstrating the visual role in constructing complex auditory cognitive maps, using a mental representation of the environment and not by direct visual information. Similarly, several studies have demonstrated that auditory perception can also be biased by haptic stimulation [Bruns et al., 2011], [Bruns and Röder, 2010]. In [Tonelli et al., 2016], the authors investigate if it is possible to use tactile stimuli despite visual ones, to modify and possibly improve auditory spatial representation, using a mental map of the environment. This hypothesis was tested using two groups of blindfolded sighted people performing an auditory bisection task. The results of these tests showed interestingly an improvement after the manipulation of the 3D model but any additional improvement after the room observation. These results show that tactile information can modulate the precision of an ongoing auditory spatial task, as the visual one does. This suggests that cognitive maps related to haptic information can participate in cross-modal calibration and model representation of space, increasing our understanding of sound propagation in space. Moreover, the absence of visual cues in infancy does not preclude that development of spatial strategies, but specific abilities might be compromised. In [Vercillo et al., 2017] it was shown, in a comparing study between blind and sighted people, how the localization of static and moving auditory stimuli is biased in blind people by an internal body-centred representation of spatial knowledge. These studies are particularly relevant to develop multimodal technologies because they inform on the way in which multisensory processing and calibration occurs, giving to computer scientists a deeper knowledge of cognitive ergonomics and usability of their prototypes. Moreover, multisensory neuroscience is particularly relevant also in education context in which the technology is introduced to enhance learning skills and possibilities of both typical children and children with sensory, motor or cognitive disabilities.

1.1.3 Auditory sensory modalities for spatial knowledge

Multisensory integration of spatial information has also some other relevant aspects to be considered related to movement ability; a recent study [Gori et al., 2012] shows that visual orientation discrimination is worse in children with movement disorder, if compared with age-matched typical children. This finding is explained proposing the cross-sensory calibration hypothesis: if the haptic sense cannot be used to manipulation, it cannot be used to estimate size. The consequence is that it cannot be used to calibrate the visual experience of size; thus, this unavailability leads to visual discrimination impairments. The idea of multisensory calibration is quite self-explained and based on a discrete number of researches. In adults, there is a general consensus on the idea that the cross-sensory information depends on how much each sensor is reliable and if there is a conflict between various signals, the perception resulting is weighted on the average of the signals. Speculate a little more on this sensors' relationship, it can be supposed that the touch is necessary to calibrate vision, because, even if haptic information does not provide knowledge about absolute scale of object size, it still has useful direct information with body-centered and body-scaled reference and that haptic knowledge is a powerful calibration tool for vision. On the other hand, primary visual cortex has specific functions connected to orientation skill, so orientation, with the head as a reference, is quite direct to be extracted. This vision dominance property calibrated the haptic sensors' information for orientation, as the haptic calibrates vision for size judgements [Burr et al., 2011], [Gori et al., 2008]. For this reason, an early vision impairment should impact specifically the haptic-orientation ability, but not the size discrimination; and, on the other side, early haptic impairments should impact the visual-size skills instead of the visual-orientation discrimination [Gori et al., 2012]. Moreover, another interesting development, for our purpose, of perceptive psychology and neuroscience has been published in [Gori et al., 2013]: in this study, the authors examine for the first time the straight connection between visual experience and auditory spatial localization in children with visual impairments. A lot of scientific literature [Lessard et al., 1998], [RoÈder et al., 1999], in fact, have demonstrated that in visual impaired population there is an enhanced auditory information processing, suggesting and underlying a compensation process between the two sensory modalities. These studies, however, have not applied a metric representation of the auditory space. A lot of efforts, in fact, was concentrated in studying pitch and timbre discrimination [Doucet et al., 2005] or in localization but of single sounds in space, like in [RoÈder et al., 1999], without requiring the comparison and the estimation of different localization in space. The sound localization ability was instead study in 2012 in sighted children, founding that in children of 6 years old the representation of auditory space is typical with slight differences with adults' skills. In non-sighted humans, however, the mentioned study, [Gori et al., 2013], found a severe impaired auditory localization ability, with no deficits on simple auditory spatial tasks and auditory temporal bisection, in agreement with the literature. The spatial auditory deficit is related to the bisection task because this kind of activity requires a space representation that remains in the memory for all the task, adding to a good topographical spatial map.

1.2 Theoretical foundation of multisensory serious-games for education

The theories that have most influenced this area of human-computer interface from an historical point of view, are Gestalt theory, cognitive psychology theories as the one on Working Memory [Baddeley, 2003], and Activity Theory, [Oviatt et al., 2017]. In addition, there are also several related theoretical frameworks that can contribute to a comprehensive state-of-art, including embodied cognition, Cognitive Load Theory and Affordance Theory. These theories rely on one side on perception-action understanding and on the other side on the concept of limited resources, both fundamental aspects of psychological and cognitive research, supported by neuroscience and multimodal human-computer interaction findings [Stein, 2012]. The large and multidisciplinary body of research on multisensory perception and multimodal interaction confirms many Gestalt, Working Memory, and Activity theory predictions that are conceptual anchors of multimodal interaction research nowadays.

1.2.1 A brief review of Gestalt Theory

In cognitive science and experimental psychology several researches in the past decades have revealed that brain processing involves multisensory perception and integration between senses [Stein, 2012].One of the major contribution, that helped this changing of paradigm from unimodal to multisensory understanding of human cognition, was the Gestalt theory, which originated in the late 1800s and that influenced nowadays understandings of objects and social perceptions. The novelty of the Gestalt school was to present a perception theory that emphasizes self-organization of perceptive experience as meaningful wholes, rather than focuses on discrete and isolated elements. In this theory, the whole is considered a qualitatively different entity than the sum of its part. The perceptual experience corresponds with underlying neural activity and this second one was an important element stressed by this school as the principle of different quality of the whole. The principle at the basis of Gestalt understanding of totality, was the discover that when elements (as lines) are combined into a whole percept (as an entire figure), the resulting properties transform qualitatively the perceptual experience and this mechanism is the fundamental of a variety of visual illusions. Research on multisensory cognition has investigated many unexpected perceptual phenomena, as the Wertheimer's demonstration in 1912 that two lines flashed successively with a precise time interval appears to move together as an illusion related to human perception of motion picture [Koffka, 1935]. The main contributions historically involved perceptual of visual-spatial phenomena, highlighted several laws and principles for perceptual grouping, as the proximity laws, symmetry, similarity, continuity, closure and others [Koffka, 1935]. The elements of the percept are grouped rapidly according to its main principles and more than one law can influence the perception at the same time. This whole percept predominance is viewed as a way on our mind to economize our mental resources, which permits a person to focus her attention to primary tasks. Although the predominance of visual-perception studies, the laws of the Gestalt school have been applied also to acoustic, haptic and other sensory modalities [Bregman, 1990]. Contemporary research on multisensory integration has clarified that there exist asymmetries, when there is more than one sensory input, on what type of signal dominates the perceptual integration. Furthermore, it also has elaborated our understanding of how spatial and temporal proximity influence the salience of the percept. In addiction to have promoted a better understanding of how it occurs multisensory perception, Gestalt theoretical principles have influenced also human-computer interaction research. For example, studies of users' multimodal way of spoken and written confirm that the multimodal constructions are qualitatively distinct from their unimodal parts. Furthermore, as we can read in [Oviatt, 2003], Gestalt theory's contributions have been applied also in multimodal communication and human-computer interactions. The Gestalt's laws predict the organizational cues between the two modalities, figuring that the common temporal dimension will organize the two modalities during the multimodal interaction. One considerable implication of these researches is the understanding that time-sensitive multimodal systems need to accurately model users' multimodal integration patterns. For this reason, user-adaptive multimodal processing is a fertile direction for human-computer interaction systems.

1.2.2 Working memory and cognitive performance in multimodal processing

Working Memory theory is one another of the more relevant contribution from the field of psychology to multimodal interfaces and interactions. The main understanding of this paradigm is that both attention and working memory are limited and constricting processing during cognitive activity. The boundary span of working memory is a limited capacity system that constraints critically the basis of cognitive functions as problem solving, inferential reasoning, language comprehension, planning, spoken and written understanding and others more. Cognitive science stressed, in last decades, the concept of limited resources of our cognitive system in theories as Working Memory Theory and Cognitive Load Theory, to address the fundamental issue of how to manage input and output modalities to optimize human performances. The working memory task is to store information temporarily in mind, usually for some seconds, without any kind of external aids, before information is confirmed and stored into long-term memory. To be consolidated, the information should be continuously rehearsal. The loss of information in working memory can depend on cognitive load, that if the task is too demanding, due to high difficulty or similar factors, can occurred. The Working Memory theory was firstly introduced over

50 years ago by Miller and colleagues [Miller et al., 1960], that described the span of working memory as built of approximately seven elements or "chunks", which can be of different types of content. The limited capacity of our short-term working memory can be expanded in some precise cases, for example if the information content involves different sensory modalities, processed in different brain regions. Baddeley and colleagues [Baddeley and Hitch, 1974] proposed a consequential theory of working memory in the 1970s, which preceded the recent neuroscientific findings on multisensory processing. According to the authors, the working memory consists in multiple and semi-independent processors, each one associated with different modalities [Baddeley, 2003]. There are specific processes for visual-spatial information, for example, such pictures or diagrams, and specific phonological process for auditory-verbal information in different brain area. These different modality-related processing systems, in the Baddeley approach, are view as highly functioning independent. In his theory there is also a central component responsible for organizing actions, directing attention to relevant information and merging information from different low-level modalities, managing the overall decision-making processes [Baddeley, 2003]. The semi-independence of the modality-specific processing is the feature that let people to use multiple modalities during a task avoiding the short-term memory limitations. An example of this feature is that if we are involved in dual tasking it is easier to maintain digits in mind if at the same time we are working on spatial task instead of a temporal one, as another numeric one is [Maehara and Saito, 2007]. The ability to expand working memory limits is particularly crucial with difficult assignments, where more information occur and need to be integrated to solve the problem. From an HCI point of view, these theoretical elements are of fundamental importance in designing computer interfaces, because they lead to two implications: (a) the human performance enhances if a computer interface supports multimodal stimulation, so that it is possible to aid complementary information processing in several brain regions at the same time; (b) multimodal interfaces that underpin processing advantages are critical if the task is difficult or the cognitive processing skills of the user are compromised. During the past decades, the research on the neural basis of memory has grown rapidly [Cools et al., 2008], confirming and contributing to understand the modality-specific brain regions and multisensory information processing, isolating the circumstances in which interference occur to obstruct the memorization process. Researches in neuroscience have confirmed that working memory is lateralized and that the right prefrontal cortex is more engaged in visual-spatial working memory, while the left more in verbal-auditory tasks [Daffner and Searl, 2008]. Working memory theory is well combined with Activity Theory, both emphasizing the dynamic processes of building memories, in a dual process of neural activation and inhibition; actively forgetting, indeed, is now generally understood to be an inhibitory neural process under conscious control, fundamental to build new memories [Anderson and Green, 2001]. To apply working memory theory to learning theory, Sweller [Sweller, 1988] introduced the concept of cognitive load. The core concept is that during learning, students can acquire schemas and let them to become spontaneous, if the methods (and the interfaces) minimize the students' attention and working memory needs, reducing the cognitive load [Oviatt, 2006]. Researchers have done a lot of studies assessing the complexity associated with instructional methods and tools separately from the intrinsic complexity of

the learning load; on this basis, the educational researchers have focused on evidence-based redesigning the learning materials and tools to help students to decrease the cognitive load and enhance the learning process. Numerous learning studies have shown that multimodal learning tools support students' learning more successfully than unimodal presentation does. In 1995, Mousavi and colleagues [Mousavi et al., 1995], for example, showed that in educational context geometry learning improves if the session includes diagrams and audio recordings, let the students to solve easily problems, comparing with students' ability to solve the same types of problems in a visual-only presentation context. In recent years, Cognitive Load Theory has been used also in computer interface design [Oviatt, 2006], supporting the research on multimodal interfaces in educational area. There are studies that demonstrate that multimodal interfaces are effective because they support students' ability to manage their working memory and reduce cognitive load, as we will see better in next paragraphs. The final implication of these research findings is that multimodal interfaces are well suited for educational application, since they involve typically higher levels of loads connected with mastering new knowledge.

1.3 Pedagogical theories on multisensory learning and embodied cognition

Educators and clinicians argue from a very long time that multisensory training can enhance learning [Fernald and Keller, 1921], [Montessori and Holmes, 1912], [Orton, 1928], [Strauss and Lehtinen, 1947]. One benefit on which all the advocates of multisensory educational methods agree is that it can engage individuals with different learning styles. According to the learning styles theories as the Kolb's or Mayer's hypotheses, some people can be "visual learners" and others "auditory learners". However, above and beyond this, multisensory is demonstratively effective at an individual level: as Treichler [Treichler, 1967] said "People generally remember 10% of what they read, 20% of what they hear, 30% of what they see, and 50% of what they see and hear". In the Montessori approach, a multisensory learning movement started approximately a century ago, most subjects use a mixture of visual, auditory, tactile and kinesthetic approaches. It is not surprising, then, that with the advent of technology in schools the interest and research in the field of multimedia (and multimodal) technology for education are growing and that are strongly enriched by perceptual research on multisensory facilitation. In 1991, with the principle of "dual coding" theory [Clark and Paivio, 1991], Clark and Paivio propose the idea that information entering the system through multiple processing channels helps circumvent the limited processing capabilities of each individual channel and, thus, greater total information can be processed when spread between multiple senses. Other researches as [Bagui, 1998] and [Harp and Mayer, 1998] show that multimodal processing reduces cognitive load because information from different modalities can be easily chunked into short-term memory and so be available to build long-term representations; in their work Harp and Mayer used this idea in different classroom studies, where animations co-presented with narrations facilitated learning of facts and concepts and long-lasting memorization. Therefore, this framework indicates key constraints on how this learning facilitation occurs [Mayer et al., 2001]. All these findings indicate that research on multisensory facilitation and perception can have important benefits in pedagogy and new educational technologies yet to be exploited. There is one more concept from philosophy and cognitive science, the embodied cognition, that can help us understand the correct theoretical framework to design multisensory and multimodal learning experience: recent research in several areas of knowledge are placing more importance on the concept of embodiment and its role in understanding human cognition and behavior, although the concept itself is not new. There is not an unique prospective of definition on the so-called "embodied cognition", as can be easily found in [Damasio and Sutherland, 1994] and [Semin and Smith, 2008], with a more linguistic approaches focused on semantics in bodily metaphors [Gibbs Jr, 2006] or more cognitive ones centered on evidence for modal and sensory representations and mental simulations [Pecher and Zwaan, 2005]. As we saw at the beginning of this paragraph, psychologists and philosophers have long considered the relevance of the role that sensory-motor interaction has with the environment for a correct cognitive development [Piaget, 1972], [Vygotsky, 1978b], [Clark and Chalmers, 1998]. Moreover, Vygotsky was one of the first and influent psychologist to highlight the relevance of tools in modelling our cognitive development and knowledge, but in recent years, with the growth of ubiquitous computing the possibility for enhancing physical environments and physical interaction have brought discussions about the embodiment of cognition and learning. During the last approximately seventy years a growing number of disciplines emphasize the role of embodiment in cognition, as we can easily found looking at philosophy [Merleau-Ponty, 1945], [Clark and Chalmers, 1998], neuroscience [Garbarini and Adenzato, 2004], [Decety and Grèzes, 2006], AI [Costa and Rocha, 2005], HCI [Winograd et al., 1986], [Dourish, 2001], linguistics [Lakoff and Johnson, 1980], and gesture studies [Goldin-Meadow et al., 2009]. In this framework, all the basis of the learning experience and meaning making is founded on sensory and motor involvement. Recent studies show the importance of embodied learning in primary schools, especially in mathematics knowledge and science [Manches and O'Malley, 2016]. In 2000, Lakoff and Núñez [Lakoff and Núñez, 2000] propose the idea that mathematical concepts ground upon bodily experiences, as tactile perception of objects and physical movement: moving in linear path, for example, can help the link to the idea of the number line. It has been showed that for the children is particularly significative from a cognitive point of view moving their bodies in a way that supports a way of thinking. In [Hostetter and Alibali, 2008] the authors suggest the interpretation of the use of simulated actions and gestures to communicate abstract ideas.

1.3.1 Flow theory

In [Csıkszentmihályi, 1975] the author introduced for the first time the concept of flow state through the study of people involved in activities such as dance or climbing. The concept of "flow" describes a state of complete absorption or engagement in an activity and refers to the optimal experience. During an optimal experience, a person is in a psychological state where he or she is so involved and focused with the goal driven activity carrying on, that nothing else seems to matter. Different past researches have shown that the flow state has a positive impact on learning - e.g. [Webster et al., 1993], and should be considered when designing learning materials. There is no doubt that the original flow activities, like hiking or playing chess, are different from the activities performed with computer and technology in general; [Finneran and Zhang, 2003] indeed have argued that activity performed in computer-mediated environments needs to be broken into the main task and the artifact used to accomplish the activity. According to the model, the likelihood of experienced flow is dependent on the interplay between the person, the task and the artifact. In computer-mediated flow studies the stages related to flow that are distinguished are: flow antecedents, flow experiences and flow consequences [Chen et al., 1999]. There has been some debate on which factors referring to flow belong in each stage, but a summary of relevant factors in each stage is here proposed [Kiili, 2005]. The flow concept is not entirely new, as it has some relevant antecedents that emphasize its importance, as the concepts of focused attention [Hoffman and Novak, 1996], the appropriate feedback [Chen et al., 1999], the perception of challenges adapted to person's skills [Chen et al., 1999], playfulness [Webster et al., 1993], the speed and ease of use idea [Skadberg and Kimmel, 2004] and other more. It is not surprising then, that the flow experience merge action and awareness, concentration, a sense of control over activity [Chen et al., 1999] and time distortion [Finneran and Zhang, 2003]. As the research states, the flow experience leads to increased learning [Skadberg and Kimmel, 2004], exploratory behavior [Webster et al., 1993], positive effect, and perceived behavioral control and for these reasons is nowadays relevant in technological designs. All these three components, in fact, person, task and artifact, should be taken into account when designing educational games. Generally, the purpose of an educational game is to provide students with challenges related to the main task so that flow experience is possible. Because of the constraints due to the limited processing capacity of working memory [Miller, 1956], all possible resources should be available for relevant information processing rather than for the usage of the artefacts. In ideal situation artefacts are transparent and allow the player to focus on the higher order tasks. Therefore, a user's prior knowledge and experiences change how the user experiences and perceives the game environment, that has to offer the learners the challenges that are in correspondence with her skills to let the experiencing flow be higher. It is also important that the challenges that a player faces in the game world is closely matched to her or his skills' level. If the challenge is significantly greater than player's skill level, she or he may feel anxiety and frustration.

1.3.2 Experiential learning

Experiential learning theories are well known by designers of digital learning environments. The framework of these theories is built upon the work of Piaget, Lewin and Dewey [Nielsen-Englyst, 2003]. One of the most influent American philosopher and pedagogue of the early Twentieth century, Dewey, is one of the primary figures associated with the philosophy of pragmatism and the functional psychology. Dewey focused his attention on the notion of experience, meaning the practical action between the subject and the environment in which she interacts. This element is important to understand his pedagogical framework, where action and experience always precede any kind of intellectualism. Experiential learning theories, therefore, consists of several models that stress the importance of direct experience and reflective observation. In adult education theory, the conception of experiential learning is an established approach and Kolb's four-stage model has a fundamental role [Kolb, 2014]. The author itself, in his work "Experiential Learning", states that Dewey, Lewin and Piaget are the founders of this approach. What is common in all the theories beyond that model is that learning begins with a concrete experience followed by a collection of data and reflective observations about that experience. In the model, a learner generalizes and form hypotheses about the experience until the final stage when the learner tests these ideas and hypotheses through active experimentation in new circumstances. What is relevant for our purpose of identifying the several aspects that are meaningful for game-based technology for multisensory learning environments is that experiential learning provides a fruitful basis for integration of gameplay; in [Gredler, 1996], for example, simulations and learning games have been categorized as tools for experiential training. A general idea of the implementation of these theories in game-learning environments can be business games, which are often based on experiential learning theories [Nielsen-Englyst, 2003]. In the next sections, we will discuss more deeply experiential gaming model.

1.3.3 Adaptive learning

Adaptive learning is a technology-based educational solution that modifies the presentation of the material basing on students' performances and preferences. The core idea at the base of this system is to capture fine-grained data and use learning analytics to enable students' profiling. The framework for adaptive learning, however, is born several decades ago, since it has its roots in cognitive psychology, beginning with the behavioral psychology of B.F. Skinner in the 1950s, and continuing through the artificial intelligence debate in the early 1970s. At that time, the idea that computers would be able to achieve human adaptability was commonly accepted, but the cost and size of computers were a barrier to its development; moreover, the user interfaces were not conducive to learning process. The first works on adaptive learning systems are usually rooted to the SCHOLAR system that was used to teach topic of geography of South America [Carbonell, 1970]. Nowadays, the interest in adaptive learning is growing; in the United States, it exists a national technology plan (NETP) that supports technology solutions, adaptive learning particularly, for their ability to teach and assess in a continuous enhancing loop. The central positive element of adaptive solution is the ability to personalize the learning experience for each user of the system. Let the students track their own learning practice means that they can develop new self-monitoring skills and be engaged actively in their education. The same feedback loop that improves students' engagement in their learning progress, it is also good for teachers. It helps them to monitor and understand each student's performance and learning path, and let them adapt their teaching on this new kind of information. They can easily identify students' difficulties and those students at risk of poor learning outcomes, providing appropriate interventions to students' needs, support and learning materials. There are many e-learning systems available nowadays, but they do not have the ability to tailor learning on students' features. An adaptive learning system has to consider:

- Many different curriculum sequences – The system should be able to change the sequencing of the learning materials in a way that makes student's learning effective, in a way such a teacher in a one-on-one session does.
- Follow the student's learning speed – Students should progress through the system only after they master the concepts previously learnt.
- Consider prior student' knowledge – Any adaptive learning system should have the skill to target student's starting point on her prior knowledge and helping her to progress toward desired learning goal. This strategy prevents frustration of struggling students and boring of other students.
- Several strategies for student engagement – Nowadays gaming has been shown to be an important way to engage students in learning. Also, frequently changing the learning materials types (video, audio, text, etc...), stimulating a multisensory integration and asking them to actively be involved in their learning session, is a strategy that can help to engage students. Adaptive learning programs that consider these elements, using strategy game simulation and serious-game, help students to maintain their attention and their first-person commitment.
- Interactive support – Rather than passively telling students what they should do next, it's important that the system is able to emulates solutions, hinting students to rethink their strategies if they are not working.
- Customized presentation – one of the most important aspect of adaptive learning system, especially for an inclusive education, is to present each learning material and lesson in the best modality for each student, following her needs. The continuous feedback loop on student's performances and behaviors should help the system to adapt the new material in the way that makes most sense to that student.

Most of the previous points relates to the ability of adaptive system to make cognitive modeling of students, in a way to understand the right cognitive flow experience and load of each student, that are two essential elements to relate correctly with students and enhance their learning. Five principles of effective gaming also serve as important elements of adaptive learning systems [Gee, 2008]:

- Sequenced challenges;
- On time and on demand information;
- Performance before competence;
- Motivation and attention;
- Timely and specific feedback.

One more element, the Zone of Proximal Development (ZPD), is a well-known concept in pedagogy and can be central in interaction design. Introduced by Vygotsky and in an ongoing definition, it referrers to the theoretical space of difference between what learners can do on their own and these activities for which they need help. Vygotsky idea was that children gradually learn by examples to solve different tasks on their own and for that reason Vygotsky and other educators believe that the role of education is to let the children experience tasks in their zone of proximal development, encouraging their individual learning [Berk and Winsler, 1995]. When the task is easy, learners can do their work on their own, in their "comfort zone". If all the work to do is always in the comfort zone, then there will be no real learning and the students will probably lose interest. On the other hand, if the work is continuously too hard, the learner becomes frustrated and will likely give up. The ideal area between comfort zone and frustration zone is where the learning will succeed, the ZPD. In this area a learner will need some help or work hard to complete her task, but she will do it. The idea of keeping the challenge appropriate, in a way that the learner is some way guided to actually do her tasks with right strategies and thinking, is widespread in the designing of adaptive learning systems. Finally, learning styles are widely used and studied as element of interest in developing adaptive learning. This term is used to refer to several competing and contested theories that aim to account differences in individual learning [Coffield et al., 2004]. According to these theories, everyone has a specific learning typology, although different theories present differing views on how the styles should be defined. The element of agreement of all these theories is that individuals differ in the way they learn [Willingham et al., 2015]. In e-learning and adaptive learning researches, often learning styles were exploited [Hwang et al., 2013], [Mainemelis et al., 2002], [Carver et al., 1999], [Graf and Kinshuk, 2007], [Graf et al., 2009], [Paredes and Rodriguez, 2004]. Learning styles theories are attempts to understand and have a defined approach to a complex issue. A closer look now reveals that one size does not fit all learners. As it turns out, doing is not always more efficient than seeing, and seeing is not always more effective than reading or listening. One of the bottlenecks to efficient learning is, as we stated in previous paragraphs, our own physiology – the way our brains has constraints on own capacity to learn and process information. It is precisely this limitation that educators must overcome through informed design of learning environments and resources. As they design lessons, create learning environments, and interact with students, they are seeking augmentations that accommodate for these human limitations. Educators are continuously redesigning learning experiences to increase and deepen learning for all students, as evidenced by the recent literature on differentiated learning [Tomlinson and McTighe, 2006]. Their efforts are much more likely to succeed when their work is informed by the latest research from the neurosciences (how the brain functions), the cognitive sciences (how people learn), and research on multimedia and multimodal designs for learning. Within those constraints, research is emerging that provides important guidelines for multimodal use of media with students.

1.4 Music and sound perception in developmental age: a powerful learning tool

Over the past decades there has been an improvement in research publications on music perception and performance and their correlates in sensory processing and generally in human brain. This increase in scientific work on music has been motivated in part by the idea that music offers a unique opportunity to investigate brain organization. Music, like language, is a complex rule-based activity, associated with a specific brain architecture. In addition, in quite all the population, sensitivity to music structure develops early in life, without a conscious effort, but – unlike language, only a minority of individuals become proficient as a musician through explicit tutoring. This distribution of acquired skills confers to music an important role in the study of brain plasticity. Differently form language, considering music it is possible to study at the same time different sensory and motor systems and processing. Given both similarities and differences between music and language, Peretz and Zatorre [Peretz and Zatorre, 2003] recognize the central role of the relationship between music and brain for cognitive neuroscience research. In the last decades different studies investigate the remarkable human brain capacity to alter in response to environmental inputs. Several studies compare adult musicians and non-musicians structural and functional differences in sensorimotor brain areas [Elbert et al., 1995], [Schlaug, 2001], [Gaser and Schlaug, 2003], auditory areas [Bermudez and Zatorre, 2005], [Gaab and Schlaug, 2003a], [Lappe et al., 2008], and multimodal integration areas [Bangert and Schlaug, 2006], [Gaser and Schlaug, 2003],[Münte et al., 2001], [Sluming et al., 2002], [Sluming et al., 2007], [Zatorre et al., 2007]. Most of the studies investigated functional brain changes and correlates of music training in childhood, but rarely it has been studied the structural brain and behavioural changes in the developing brain as a response to long-term music training to specifically investigate whether structural brain differences seen in adults are product of "nature" or "nurture". The positive aspect of this kind of study is to examine also cognitive and behavioural changes in parallel with brain changes in response to music training. An important role in the developing research on multimodal integration and technology for learning, is covered by Marc Leman's embodied music cognition theory [Leman, 2007]). In Leman' s theory, the core idea is that cognition is more than just a mental processing. Moreover, gesturing is not an isolating aid to thinking through a problem, but it is a relevant part of the cognitive process itself. Cognitive processes are, then, more than just mental processing and its brain correlates and are fundamentally "embodied". The focus of embodied cognition is not far from the one of grounded [Barsalou, 2008] and extended [Clark, 2008] cognition. The main difference is that embodied cognition thereby emphasizes the role of the body as a mediator in experience understanding. Furthermore, music cognition is situated (embedded) in the environment, and enacted, let's say realize through action [Barsalou, 2008]. The role of the body is particular, because it casually connects with a physical environment and with an experience of the subject [Merleau-Ponty, 1945]. The physical environment can be described objectively, but the subject's experience can be described only in a subjective way. In the same way, musical gestures from an objective point of view are body movements, but they have also an important experiential component related to intentions, goals and expressions. The core idea is that there is an intentional level in music interaction that is established through body movements and imitation of physical information provided by the musical environment. If the body acts as a mediator between the environment and the experience, it will build up a repertoire of gestures and consequent actions. This repertoire creates a so-called gesture/action-oriented ontology, that comprises the set of memorized actions and perceived outcomes useful for the next action and perceptual interpretation. Music gestures in Leman's interpretation can be fully included in this repertoire. An interesting feature of music gestures is that this concept applies well both to sounds and body movements [Godøy and Leman, 2010]. Research investigates whether it is possible to correlate sonic gestures (i.e. identified in music) sound-producing gestures (i.e. hands movements), and that these gestures can be understood as concatenations of more elementary gestural components. The constraints to this relation are defined by physical a physiological limitation as well as by cultural elements as preferred movements. In [Desmet et al., 2012], the authors analysed clarinet playing gestures in relation to musical expressiveness and intentionality of player. A theoretical contribution declined by embodied theories is the entrainment concept, that considers how resonant systems adapt to each other in synchronization event. Often this concept is related to timing aspects only in relation to synchronization, but recent studies consider the idea that gestures may condition entrainment, moving the focus of entrainment to include spatio temporal dimension rooted by bodily expressions. Related to this concept, in 2010 Leman and Naveda [Naveda and Leman, 2010] elaborated the concept of gestural topology of point cloud, in which the idea is that with gross motor gestures, there is a large spatial region in which the body parts can be at a particular point in time, which permits synchronisation and flexibility in entrainment. On the other hand, fine motor gestures, such those used to play a musical piece, required a high spatio temporal deployment. In this case, the tolerance for synchronization variability is lower and so it is the flexibility of entrainment. A simple and clear example is music rehearsal, that often rotate around the needs of synchronization between players and between the several music voices. The relevant element of music embodied theory is that this gesture/action repertoire forms a more complex mechanism that controls the interaction between environment and personal experience. This mechanism, the action-perception coupling, is a hot concept to master in recent cognitive science [Wolpert et al., 2011]. In [Leman, 2012], the author claims that there are at least two mechanisms to consider in learning, the which he called "sensorimotor loop" and the "action-perception loop". The first one is a low-level mechanic where motor gestures are driven by sensory input from the environment. On the other hand, the action-perception loop is a high-level loop that involves gestures repertoire. The author explains this concept imagining a clarinet performance: the sensorimotor loop is the one that let the player to maintain the control of the mouth and breath according to the sound she is producing, while the action-perception loop can be trigged by searching patterns of fingering positions, for example as a controller for the expressive production of the sound. The assumption is that both action and perception can be controlled by an action-perception processing system, that involves different loops. The prediction principle works also for the listener: when a person perceives a sound, s/he can rely on previously learned knowledge to understand the action for the sound. This principle is fundamental to understand why action is embodied and it is also necessary to modulate in a correct way our multimodal interaction with music. Typically, with everyday sounds, people react on how the sound is produced. On the other hand, if the sound produced is abstract, people tend to react to features of sound that can be reproduced with body movements [Leman, 2012]. The paradigms of embodied cognition and embodied music cognition are interesting for developing educational multimodal applications, because introduce viewpoints on gestures, movements and on the relation between sounds and the experiences mediated by the body. This framework provides an approach to deal with the experiential component of music experience. Embodiment assumes the existence of mirroring processes that facilitate both the encoding of gestures into sounds and vice versa the decoding of sounds in expressive gestures. Moreover, mirroring implies underlying mechanisms, that work in meaning formation processes, rooted in sensorimotor principles [Leman and Maes, 2015]. Empirical studies focus on the role of body and non-corporeal technological mediators [Nijs and Leman, 2014] and the role of expressive gestures both in music playing and listening [Van Dyck et al., 2014]. In contemporary research there are, indeed, two separated and well-defined trends: the first one is to broaden the perspective on embodied music cognition, linked it to affective research [Cochrane et al., 2013], music engineering [Kirke and Miranda, 2013] and brain studies [Arbib, 2013], [Maes et al., 2014]. However, there is also a trend in narrowing the embodied perspectives, because the evidence of embodied cognition seems to be weaker in other domains. Several studies introduce new movement analysis methods, enlighten the strong relation between music perception and body behaviours [Sievers et al., 2013]. The chance for developing technology is to be able to use this cognitive framework to apply the control of such complex phenomenon in applications for education, rehabilitation, art and other human domains. Moreover, embedded cognition theories have also an impact on pedagogical research. A work on gestures and their role in maths learning is [Radford, 2009]. In this paper, the author considers, starting from embodied cognition theory and Gehlen "sensuous cognition"' concept, some first evidences on the multisensory components and the role of body movements and gestures in mathematical thinking of primary school children. On the same direction, Novack and colleagues [Novack et al., 2014] in 2014 published an interesting research in which, considering previous results on gestures relevance for maths instruction [James, 2010], [Kontra et al., 2012], [Sommerville and Woodward, 2010], [Wilson, 2002], investigate whether gesturing promotes learning because it is itself a physical action, or because it uses physical action to represent abstract ideas. Their first results show that gesturing itself promotes a better learning, then directly acting on object, suggesting in this way that probably gesturing's benefits for learning may involve different features that differentiate it from action [An and Tillman, 2014].

1.4.1 Music attitude developing: nature or nurture?

The development of music competences is a subject that, more than others, divides opinions. There are, indeed, even in pedagogy practices, people that consider music has something for which particular schools and *talent* are needed. Otherwise, there are researchers that sustain that music perception does not need particular learning path, not more than the perception through the other sensory modalities or the language itself need. Of course, great part of this discussion depends on music comprehension definition. Most of the people without a specific music training are able to appreciate music, sing a simple melody, recognize a good performance, understanding mistakes (especially for intonation parameter) that may occur and these are all competences resulted by an implicit learning. Listening to a piece of music requires the processing of an enormous amount of information in a very fast and essentially automatic way. This process is completely comparable to what happens when listening to a speech in a known language. This type of automatism is often interpreted as the result of an innate brain mechanism. However, even these innate mechanisms also need environmental support, i.e. perceptual or musical learning that takes place through exposure to musical stimuli typical of one's own culture [Schön et al., 2007]. Moreover, music reading and instrument playing requires specific learning path and time to be mastered by the students. An interesting work investigating the structural and behavioral brain changes in developing brain due to musical training is [Hyde et al., 2009]. In this research, the authors investigated structural brain changes in relation with behavioural ones in children (mean age at start of study 6.32 years old) who received 15 months of instrumental music training, compared with a control group who did not. The positive impact of this paper is due to the attempt of authors to examine both brain and behavioural changes in the development to understand if the brain differences in adult musicians are due to their training or to genetically-driven elements. Structural changes in motor and auditory areas (brain regions of critical importance in music training) were correlated with behavioural important improvements in motor and auditory-musical tests. Even if the results could not completely rule out the hypothesis of pre-existing biological predictors genetically determined for musicians, these findings support the hypothesis that brain differences seen in adult musicians relative to non-musicians are more likely to be the product of intensive music training [Schlaug et al., 2005]. Children engaged in the musical training, showed greater improvements in fine-motor ability (tested by fingers dexterity in both hands) and in auditory melodic and rhythmic discrimination skills. Contrary to [Schellenberg, 2004], [Vaughn, 2000] children who studied music instrument for 15 months did not show a superior improvement in visual-spatial and verbal transfer domain outcomes than children who did not receive musical training. One of the most interesting results the authors published is the high deformation difference in the left posterior pericingulate region. This region lies near to Brodmann area 31, in the transition area between posterior cingulate and occipital cortex and it is involved in the integration of sensory (mostly visual) information and limbic system. Such integration is involved in learning to read music notation, for example, and related music and its emotional content. Overall, these findings underlie that plasticity can occur in brain regions that control primary functions to play a musical instrument, but also in brain regions that might be responsible for the multimodal sensorimotor integration underlie instrumental music learning. Nevertheless, it has been proposed one cognitive theory of musical development that looks at rules and representations of specifically *musical* understanding in children, namely *The music learning theory* of Edwin Gordon [Gordon, 2003]. The core aspect of Gordon's theory is to consider both experience and training in shaping children's music thinking. A central concept in Gordon's theory is, indeed, audiation:

> Audiation is the process of assimilating and comprehending (not simply rehearing) music heard in the immediate past or days, weeks, months, or years ago. Musicians *audiate* also when assimilating and comprehending in their minds music that may or may not have been heard but is read in notation, composed, or improvised. Aural perception takes place when sound is reacted to without comprehension the moment it is produced. Sound is audiated only after it is aurally perceived. In aural perception, persons are dealing with immediate sound events. In audiation they are attending to delayed musical events.

Gordon claimed that ”audiation is to music what thought is to language” [Gordon, 2007]. He identified four core items of music, namely listening, speaking, reading and writing and suggested that individuals' aptitudes and progress determined the extent to which these items are acquired and used by the subject. For this reason, in young children or generally people without any kind of music learning, he proposed three types of *preparatory audiation*: (a) acculturation, (b) imitation and (c) assimilation, which were held to prepare them to formal music instructions, developing their audiation. According to the theory, there are eight types and six stages of audiation; each stage follows a developmental sequences, whereas the eight types do not, but some types serve as a preparation for other. He constructed a complex and detailed account of music learning sequences on the basis of these foundations, comprehending three elaborated sequences dealing with skills, tonal and rhythm content. A great deal of Gordon's empirical research work was devoted to the development of psychometric test designed to describe, measure and evaluate musical aptitude achievement: his Musical Aptitude Profile [Gordon, 1965], for example, is still one of the most used test of musical aptitude internationally. In a review, Colwell and Abrahams [Colwell and Abrahams, 1991] argued that there was little support from research other than that which has been directly collected by Gordon's students and co-workers. Moreover, Swanwick and Runfola [Swanwick and Runfola, 2002] also pointed out that the theory may be in some extent ethnocentric, considering that all the learning sequences in the taxonomy are specifically based on Western classical music. Gordon's theory deal specifically with developmental progress in that they are based on developmental psychology underlying age-related changes in musical behaviour.

1.4.2 The natural link between music and mathematics

Children, from very early age, learn through being active and interactive with the environment. They use all their senses to learn about the world around them, connecting new information to what they have already learnt. It is, indeed, a creativity process and for such reason this aspect is highly important for a successful learning process [for Education, 2006]. Introducing music in educational technology context, is one of the way to stimulate creative thinking. In [Gerson et al., 2015], the authors examined the role of active experience on sensitivity to multisensory synchrony in six-months-old infants in a musical context. The results showed that active experience provided a unique benefit contrary to observational one, giving insights on the embodied roots of early music perception and cognition. As mentioned above, music has been deeply studied in the last decades from a cognitive and neuroscientific prospective. Research in music therapy suggests that maths and music are strictly interconnected in the brain since early age [Burack, 2005], and music is related to very primal parts of the brain and its development [Hudson, 2011]. This implies that even the youngest children have the potential to inherently respond to the music and the mathematical patterns it contains. In 2008 Spelke [Spelke, 2008] proposed that the experience the children made of operating melodies, harmonies, and rhythms may activate portions of brain, facilitating children's ability to learn geometry issues, such as representation of numbers and estimation. The interest in the connection between music and cognitive and strategic thinking abilities led in 1993 Rauscher and colleagues to present an empirical study on causal relationship between listening and spatial-temporal reasoning [Rauscher et al., 1993]. The study presented music listening assignments, as independent variable for participants, and spatial-reasoning tasks, assessed using the sub-tests from the Stanford-Binet Intelligence Scale. The published results showed that participants that listen to a specific Mozart's composition had significantly higher spatial-reasoning scores than those who were assigned to the group listening to relaxing music or to the control group listening to silence. This landmark study was indeed quite controversial; a variety of study were afterwards undertaken to study the "Mozart effect", with different results and a great variation on the theme of researching how music may impact mathematical abilities. Within these studies, different types of music were explored, using various instruments, quantitatively and qualitatively different, to assess mathematical abilities in several tasks [Rauscher et al., 1995], [Ivanov and Geake, 2003], [Hui, 2006]. This great variety of study and interpretation of Mozart effect's concept lead to misunderstanding in later researches and results so that in 2006 Rauscher and Hinton [Rauscher and Hinton, 2006] clarify that the improvement in spatial-temporal skills associated with music that they have found is related to music listening, in a specific target population of adults (college students) and that the improvement is for up to 15 minutes. While the scientific interest in music cognition and its relationship with learning, mathematical thinking and its effects on the brain grew, researchers have also examined the relationship between learning to play a music instrument and students' mathematical abilities. Results have shown a strong positive relationship between music learning and mathematical thinking skills [Kafer and Kennell, 1999]. Several studies have consistently found the same positive results in all K-12 grades students including preK-K [Rauscher and Zupan, 2000], [Lahav and Mioduser, 2001], [Whitehead, 2002], [Cox and Stephens, 2006]. By integrating music and mathematics, teachers can help children to achieve learning standards in the second subject and make easy mathematical connections in context not strictly related to mathematics. Rauscher and colleagues [Rauscher et al., 1997] found that music promotes the development of such thinking skills as recognizing patterns and logic strategic thinking. The idea of patterns is a powerful tool that can be used to link mathematics and music to every other curriculum area. Work with patterns enhance children's thinking and reasoning skills, because it requires analyzing the pattern and find a rule, be able to communicate it and make a prediction on the outcome [Edelson and Johnson, 2003]. Musical elements such as steady beat, rhythm, melody and tempo, indeed, possess inherent mathematical principles such as spatial properties, sequencing, counting, number line, patterning, and one-to-one correspondence. Neuroscience research indicates that steady beat affects attention behaviors in humans, since we process it in the premotor cortex of the brain, a cerebral area related to attention [Bengtsson et al., 2009]. In 2010 Zentner and Eerola [Zentner and Eerola, 2010] found that children, aged 5-24 months old, were more engaged in rhythm-only stimuli that with speech-only stimuli. This study indicates that children can potentially be more engaged when listening a beat instead of verbal-only instructions. Therefore, listening to a steady beat pattern during mathematics learning activities in early childhood could promote better attention and engagement. In their research Zentner and Eerola suggest that infants and toddlers have an innate capability to not only see patterns but also hear them in music. Reinforcing this capability by teaching patterns through music may benefit children's cognitive skills [Meltzoff et al., 2009], but it is also a key for the concept of emergent mathematics, which parallels the idea of emergent literacy [Geist et al., 2012]. As for literacy, emergent mathematics suggest in fact that mathematics learning begins very early in life; it is related to many other developmental skills; it develops in real-time situations in which the child is actively engaged. Nowadays there are many good hands-on patterning materials available for teachers, but they tend to be mostly based visual/spatial information processing, such as colored blocks or tiles. However, as we mentioned, neuroscientific research highlights the need of multisensory information, since every child use a preferential sensory modality for specific knowledge. Furthermore, one of the first patterning experiences that children have in infancy is based on auditory and haptic senses [Meltzoff et al., 2009]. Infants do experience of steady beats and rhythmic patterning from the very first moments of their life through their parents' lullabies, used to soothe or rock them to sleep. These songs contain many complex patterns. The pattern is processed in several brain regions as the child listens to, feels, watch and finally internalize the pattern. This early exposure to patterns is not, of course, strictly connected to teaching mathematics content, although the caregiver is introducing the first blocks of mathematical understanding [Clements et al., 2011], [VanDerHeyden et al., 2011]. These and other patterning and mathematical experiences are naturally part of infants' everyday experiences and can support the future learning of mathematics, literacy and other formal subjects. However, studies suggest that, especially in the early years, the most appropriate way to promote appropriate learning is through play and other everyday activities [Phillips-Silver and Trainor, 2005]. Patterning is a key benchmark for the National Council of Teachers of Mathematics (NCTM) and part of many states' preschool learning standards [Papic et al., 2011]. Since 1998, the Eisenhower Southwest Consortium for the Improvement of Mathematics and Science Teaching (Eisenhower Southwest Consortium for the Improvement of Mathematics and Science Teaching, 1998) published an interesting report in which the authors presented the positive use of music, body music and dance to enhance the learning ability of children in two different schools in New Mexico. Music plays an important role in patterning experiences for children; the relationship between music and mathematics can be, indeed, considered as the combination of the two subjects in an interdisciplinary framework and as an overlap between the cognitive processes and individual abilities during both musical and mathematics experiences. Musical elements as melody, rhythm, scales, intervals are related to several objectives of K-12 mathematical teaching such as numbers line and operation, geometry and data analysis [Fauvel et al., 2006], [Harkleroad, 2006]. Activity as music composition and musical instrument design can be used to address several mathematical topics [An and Tillman, 2014]. Music activities and materials are a good resource to promote emergent mathematics [Geist et al., 2012], [Southgate and Roscigno, 2009], considering that music keeps children engaged in mathematical activities for long time, promoting a positive attitude toward mathematics and support the construction of concepts in an appropriate way. Moreover, as we introduced rapidly in first paragraphs, music and language are strictly connected and, in last years, music has been used as a tool to increase learning experience also in this area. In [Salcedo, 2010] the author investigates whether English native-speaker students learning a foreign language can benefit from integrating music into their activity. The measures done included text recall after listening a songs and text passages; delayed recall and involuntary mental rehearsal. Results showed that text recall was better in the music condition that in the text passage. Song condition did not influence the delayed recall, but impacted involuntary the mental rehearsal. It may be argued that a causal link may be influenced by other environmental factors, but there are several conceptual similarities, included the already analyzed patterns, making connection and recognizing relationships, working with numbers, fractions, shapes, measures, counting, sorting and matching. Music-making and improvisation methods contain many arithmetical concepts, for example time signature, denoting length and division of bars as element of a fraction, counting notes and bars and division in rhythmic patterns. Rhythms are often represented in several ways in addition to western musical notation, which represent geometrical or arithmetical concepts, for example with a "clock-based" spatial representations. In [Courey et al., 2012], the authors examined the effects of an academic music intervention on conceptual understanding of music notation, fraction symbols and size, equivalence of third grade students. In their study, students were assigned by class to their general mathematical lessons or to receive academic music instruction two times a week in 45 minutes lesson for 6 weeks. Academic music students used their conceptual understanding of music and fraction concepts to inform their solutions to mathematics problems. Fractions, indeed, are one of the most difficult concept to master in elementary curriculum [Cramer et al., 2002]. For many students, this first difficulty in understanding fractions continues to the other school grades, delaying and preventing development of mathematical reasoning [Mazzocco and Devlin, 2008]. The heart of mathematical learning is symbolic representation of an actual quantity. This symbolic representation consists in signs and rules with intentional characteristics and goals, that novice learners may not be ready to perceive. Influenced largely by Vygotsky [Vygotsky, 1978a], mathematical development has been considered as a learning tools, elements of specific communicative systems, different by culture. In semiotic perspectives, symbolizing involves manipulations with cultural tools such as physical objects, pictures, gestures, computer graphics, verbal expressions and more others. However, these tools do not convey meaning, but they are just channel through which learners can construct meanings [Chandler, 1994]. Using a semiotic approach to mathematical teaching, allows the teacher to use gestures and manipulatives, choosing symbols that encourage students to construct and understand concepts. The core idea is to arrive at meaning through active interactions and activities that link together several semiotic representations, to help the learner to organize actions and thoughts in space and time [Radford, 2003]. The authors of Academic music research used this conceptual framework to design a series of activities to engage students in visual, speaking and gesturing actions, in which students can progressively experience fraction size and equivalency. To ensure experiential multimodal learning, authors design their instructional intervention, using components of Kodaly system of music education. The first six lessons focused on music notation and temporal values of notes in four fourth times. The last six lessons focused on connecting proportional values of music notes to other signs or fractions representations and then to formal mathematical symbols. This method was designed to introduce beginning fractions concepts to third-grade students with and without learning differences. Results showed an improvement in fractions understanding and correct use in proper contexts in students who received Academic music training compared to the control group, underlying the utility to use music knowledge to easily learn mathematical notions. Musical activities are naturally embodied through activities such as clapping or tapping [Repp, 2005] or dancing [Juntunen and Hyvönen, 2004]. For example, in a study of 2011 [Fischer et al., 2011] the positive influence of dance and movements was investigated in a magnitude comparison task, using a digital dance mat that allowed bodily responses as input device. There are several demonstrations of the benefits to linking mathematics teaching and arts. As we saw, studies of interdisciplinary music and maths lessons found positive effects on students' mathematical abilities and there have been found classroom strategies exploring the using of clapping and chant to teach addition, songs to teach subtraction, note values for fractions, counting along the number line using percussion instruments and fractions using measures of music [An and Capraro, 2011, An and Tillman, 2014, An et al., 2011]. In 2015 An and Tillman [An and Tillman, 2015] proposed an experiment on 56 elementary students, with a random assignment pre-test – post-test control group design, used to examine students' changes in mathematical ability. The results demonstrated that music-mathematics integrated lessons had statistically significantly improve the music group students' mathematical abilities.

1.4.3 Music and Dyslexia

Music shares several basic processes with other human activities, and this can be particularly evident comparing speech and music [Chobert et al., 2011]. Both relies on sound processing and require a precise representation of several sound features, such as timbre, pitch, duration and so on. In making music, the attention on all sound features and quality is higher, because sound quality is what a musician works on from the very start. This may also explain why music training enhances processing of sounds features that play also a role in speech processing [Kraus and Chandrasekaran, 2010]. It has been showed, in fact, that adult musicians have a better representation of speech sound features both in terms of pitch and formants [Wong et al., 2007]. One of the relevant properties of music is structuring sounds in time and in a tonal space, and for this reason is not surprising that music-dependent brain plasticity goes beyond subcortical and primary auditory and sensorimotor cortex, affecting more integrated functions. There is for that evidence that music training facilitates language learning. In 2008, Forgeard and colleagues [Forgeard et al., 2008] found a strong relationship between musical discrimination abilities and language-related skills. In normal-reading children, indeed, musical discrimination abilities predict phonological and reading skills. This relationship is stronger, as expected, in children with music training than in the control group without specific music knowledge. In dyslexic children the musical discrimination task predicts phonological skills, which in turn predicts reading abilities. Furthermore, normal-reading children with music training have better performance in melodic discrimination than both normal-reading children without previous specific music training and dyslexic children. Rhythm organizes music event and perception into patterns and rhythm perception in music is studying using metric tasks. However, metric tasks play also a relevant role in organizational structure in phonology, in which speech prosody has a similar role to rhythm in music, structuring patterns and forms too. Both music and speech unfold in time, so understand the rhythm (or period in time) in which strong and weak accents (beats) occur is a critic ability to meaning both music and speech. In music, the "musical syntax" is due by both rhythm and pitch of different notes in the overall sequence [tha, 2005], whereas in language the same role is covered by prosodic structures, that has, indeed, been described as "phonological grammar" [Port, 2003]. As we saw, there is an evidence in rhythmic perceptual difficulties in dyslexic children. Development inefficiencies in basic auditory information processing might affect both language and music abilities development. Additionally, in [Huss et al., 2011] the authors explore the relation between musical metric perception, auditory perception of an envelope structure and phonological awareness (PA) with reading in both typical-developing children and children with developmental dyslexia, founding a relation between music and metric perception and PA and envelope structure and PA. Rhythm in music expresses at least two components of temporal organization, the periodicity (namely the metric structure of the phrase), and the patterning of events that can be grouped into similarly-structured groups, the phrase structure. In language, speech rhythm has a similar organizational role. The periodicity in speech signal is not perfect, but it is considered more in terms of motor constraints inherent in producing syllables [Chandrasekaran et al., 2009]. Results show that individual differences in musical rhythmic perception and perception in amplitude of envelope structures are connected and that music metric sensitivity predicts phonological awareness and reading development. These results confirm that difficulties in metric processing are associated with basic auditory time processing difficulties, suggesting a primary sensory impairment in developing dyslexia. More recently, in 2014, Flaugnacco and colleagues [Flaugnacco et al., 2014] showed a strong correlation between several temporal skills and phonological and reading abilities, confirming previous studies. Furthermore, in 2013 Tiernay and colleagues [Tierney and Kraus, 2013] show that the variability in tapping to a beat correlates with reading and attentional tests performances. Moreover, six months of piano lessons are enough to improve spatial-temporal reasoning in preschool children [Graziano et al., 1999], and piano lessons also enhance vocabulary and verbal sequencing skills in primary grade students [Piro and Ortiz, 2009]. These researches are confirmed by recent study on improvement of reading and related disorders with a musical training [Habib et al., 2016]. In this article Habib and colleagues tested the efficacy of a specially-design Cognitive-Musical Training (CMT) method. The researchers' proposal is based on three principles:

- Music language analogies: based on neuroscientific evidence on commonalities between music processing brain regions and language brain regions, training dyslexic children with music can help them to enhance their language processing skills;
- The "pattern" characteristics of music, that we introduced above, and the temporal and rhythmic features, could have a positive effect on the several dimensions of the temporal deficits that some types of dyslexic have;
- The last principle is the cross-modal integration, based on evidence of the impaired connectivity between brain regions in patients with dyslexia and related disorders. On this based, the authors proposed a series of musical exercises involving jointly and simultaneously multisensory perception (visual, auditory and somatosensory systems), and motor systems, considering particularly rhythmic perception and production and an intensive training on several auditory features. One of the most important achievements is that most improvements persisted after an untrained period of 6 weeks, providing important arguments in favor of using musical rehabilitative materials to support children with dyslexia. Showing that music and language share several sensory and cognitive processes and that music training can enhance language abilities [Moreno et al., 2008], it is not surprising that several researches hypothesize that music training may be effective in rehabilitation of motor and cognitive disorders in different clinical populations [Tallal and Gaab, 2006], [Besson et al., 2007], [Särkämö et al., 2008], [Schön et al., 2008], [Altenmuller et al., 2009], [Kraus and Chandrasekaran, 2010], [Goswami, 2011], [Patel, 2011], [Amengual et al., 2013].

Chapter 2

Multimodal technologies for education

2.1 Multimodal user interface: a review

Multimodal systems are computer systems endowed with multimodal capabilities for human-machine interaction and able to interpret information from various sensory and communication channels. One of most relevant acquisition of it, that makes it easy to understand why these solutions are so widespread, is the usability of multimodal interfaces, that support the users' ability to select input modalities and shift among them during the interaction. Another important aspect of this technology, especially if related with the increasingly use of smartphone devices, is flexibility of multimodal-multisensory interfaces that support multi-functional use. The adaptability of these platforms is something interesting, in light of developing new solutions for educational. Recent findings have revealed that more expressively powerful interfaces can stimulate cognition in a more effective way than the keyboard and mouse interfaces; the idea is that the multimodal paradigm can substantially improve the support for human cognition and performance. From a communications point of view, it can be argued that more expressively powerful input modalities, and multimodal combinations of them, can directly facilitate our ability to think clearly and perform well on tasks. Moreover, they also support the cognitive and performance enhance, because they minimize own cognitive load. The reducing of working memory load is possible using multimodal input, as combining speech and writing for example, because their average length is reduced by conveying spatial information with gesturing and non-verbal communication [Oviatt and Cohen, 2015].

Multimodal systems have developed rapidly during the past decade, with steady progress toward building more general systems, as well as more transparent human interfaces than ever before [Benoit et al., 2000], [Oviatt and Cohen, 2000]. Major developments have occurred in the hardware and software needed to support multimodal systems, as well as in techniques for integrating parallel input streams. Multimodal systems also have diversified to include new modality

Figure 2.1: Differences between MUI and GUI user interface from [Dumas et al., 2009]

GUI	MUI
Single input stream	Multiple input streams
Atomic, deterministic	Continuous, probabilistic
Sequential processing	Parallel processing
Centralized architectures	Distributed & time-sensitive architectures

combinations, including speech and pen input, speech and lip movements, speech and manual gesturing, and gaze tracking and manual input [Cohen et al., 1997, Hennecke et al., 1995, Turk and Robertson, 2000, Zhai et al., 1999]. In addition, the array of multimodal applications has expanded extremely rapidly in recent years [Cohen and McGee, 2004, Nock et al., 2004, McGee and Cohen, 2003, Neti et al., 2000, Oviatt et al., 2003, Oviatt et al., 2004, Oviatt et al., 2005, Oviatt and Cohen, 2000, Pankanti et al., 2000, Reithinger et al., 2003]. From [Dumas et al., 2009], we insert here an effective table to underlie the differences between a standard graphical user interface (GUI) and a multimodal one (MUI):

In standard WIMP interfaces (Window, Icon and Menu for example), there is a singular physical input device, used to control the position of a cursor, and the information is organized in windows and represented with icons. Whereas, in multimodal interfaces, various modalities can be used as input streams (voice, gestures, body movements, etc.). Due to the multiple recognizer necessary to interpret multimodal input and the continuous property of input streams, multimodal systems depend on time synchronized parallel processing. Further, the time sensitivity of multimodal systems is crucial to determining the order of processing multimodal commands and maintain the real-time processing, when needed. Multimodal systems can be also very resource demanding in some cases (e.g., speech/gesture recognition, machine-learning augmented integration). The advantages of these interface design are elucidated in the theory of cognitive psychology, as well as human-computer interaction studies, most specifically in cognitive load theory, gestalt theory, and Baddeley's model of working memory, as we have seen previously. Reeves et al. [Dumas et al., 2009] defined the following "guidelines for multimodal user interface design":

- Multimodal systems should be designed for the broadest range of users and contexts of use, since the availability of multiple modalities supports flexibility. For example, the same user may benefit from speech input in a car, but pen input in a noisy environment.
- Multimodal systems should be designed for the broadest range of users and contexts of use, since the availability of multiple modalities supports flexibility. For example, the same user may benefit from speech input in a car, but pen input in a noisy environment.
- Designers should take care to address privacy and security issues when creating multimodal systems: speech, for example, should not be used as a modality to convey private or

personal information in public contexts.

- Modalities should be integrated in a manner compatible with user preferences and capabilities, for example, combining complementary audio and visual modes that users can co-process more easily.

- Multimodal systems should be designed to adapt easily to different contexts, user profiles and application needs.

- Error prevention and handling is a major advantage of multimodal interface design, for both user- and system-centred reasons. Specific guidelines include integrating complementary modalities to improve system robustness, and giving users better control over modality selection, so they can avoid errors.

The design of computer interfaces has historically been a competence of engineers. As Oviatt stated in her works, in the last years it has become essential to understand interface design from a deeper cognitive science and communication processing point of views, especially to support educational application. The design of adaptive and more effective educational interfaces, indeed, requires adopting a human-centered perspective, rather than a technology one. The previous development of computer as a technology-centric activity has been based on the belief that users can and will adapt to technology applications, whatever they are, with the needed training. This belief implies the expectation that users can adapt their way of interaction to accommodate to the system's processing abilities. One problem related to this view is the time that is needed to teach users how to interact fluently with the technology, that eliminates time for another subjects' learning. The other problem is the gap between the high performers and lower-performing students that experience greater cognitive load once that technology is introduces, which cause the expanding of the achievement gap [Oviatt and Cohen, 2010]. From a theoretical point of view, educational interfaces that enhance communication can maximize students' effort associated with the construction of new schemas. The authors in their research investigate students' performance on biology problems, using a table, a hard-copy paper and pencil and a digital paper and pencil. Despite the expectation and the fact that students believed that tablet interface provides the best support to their performance, the only interface that significantly enhance students' performance is the digital paper and pen. In the paper, the author discussed the implication of this result, as an example of how a multimodal device can be optimized to support concretely learning and improve the awareness of the impact of digital tools on users' performance. A human-centered design, instead, promises to model its approach on users' behavior. This framework includes modeling constraints based on humans' ability to attend and learn, to perform and interact, so that interfaces can be designed in a more intuitive, easier to learn and to use way. The potential impact of this approach is to design lower load and more effective interfaces, needed to support lowest performing students and lifelong learning. In HCI, multimodal techniques can be used to create different types of interaction. As we previously seen, particularly interesting are perceptual, attentive and enactive interfaces, especially if we are seeking for multimodal educational experience. Perceptual interfaces are multimodal interfaces highly interactive [Turk and Kölsch, 2004], characterized by a natural and efficient human-computer interaction. They seek to use sensing and rendering technologies respectively as input and output to provide interactions not feasible otherwise with WIMP and common I/O devices (such as keyboard, mouse, monitor), with a computer vision dominance in many cases. Attentive interfaces are context-aware and rely on person's attention as primary input [Selker, 2004], using gathered information to estimate the best time and approach to communicate with the user [Oulasvirta and Salovaara, 2004].

2.2 Multimodal technologies for learning

A recent and interesting example of multimodal application for learning, it is represented by "Block Magic" [di Ferdinando et al., 2015], a multimodal project developed in the framework of LLP-Comenius programme. The project took its moves from the understanding that technology can useful contribute to learning in infant and primary schools. Block Magic considered the haptic experience of children as crucial in primary educational contexts. The targeting audience of the project wre children aged 2.5 to 7 years old.The main core of this work was to help children to learn autonomously, teaching a broad range of specific logical and strategic thinking, mathematical, language and social skills. The actual manipulative educational materials, like logic Blocks, already used at schools, have the strong constraints to require a constant adults' supervision and to let few children at time to explore and use these materials for learning tasks, while the introduction of technology smart objects has the goal to open a new learning and teaching environments to enhance neuro-cognitive development especially in training context. Touching a block or a sequence of blocks with a wand or a special glove, the system generated a feedback for the child (e.g. by talking) that changed from one activity to the next or at a different level of the same task. The project was based on RFID technology to reduce the cost and increase the usability for non-technology expert users. This approach was founded on the comprehensive idea of "disappearing computer", that contribute to innovative ICT and multimodal solutions. A relevant field in which multimodal technology has been widely developed in the last years is the cultural heritage. Cultural expression, in fact, includes also intangible live expressions, which involve knowledge and skills such as music, dance and human skills. These cultural materials are traditionally transmitted orally or by gestures and are modified over a period of time due to collective re-interpretations and recreations. In this area, multimodal technology can be successfully employed to maintain and contribute of this heritage, improve and widespread its presentation and provide universal access to cultural resources, supporting cultural and educational inclusion. i-Treasures [Cozzani et al., 2017] was a project centered on the development of multimodal technology for cultural heritage diffusion, education and maintenance. There already were other projects in which technologies were adopted for supporting teaching and learning and rarely for cultural expressions [Dania et al., 2011], [Bauer and Dammers, 2016], but none of these studies involved learners in practicing one artistic expression. The innovation presented by i-Treasures consisted in the fact that sensors and full-body movement technology were used to collect, analyze and model the complexity of performances with a cultural and educational purposes [Pozzi et al., 2015]. A previous work with a similar conceptual framework was i-Maestro [Ng and Nesi, 2008], where sensors were used to teach and learn musical skills. In that project, the combination of different tools, as for motion detection and analysis as for music annotation, let to realize a multimodal environment in which music annotation of a score with bowing symbol is in real-time with the violin execution, with pedagogical and inclusive positive effects.

New learning experience may help students or learners through offering them information using channels and methods easier to understand, personalized, trying to meet the different students' needs. Those alternatives can be particularly relevant for children and people with special needs. In [Nusir et al., 2013], the authors investigated the impact and effectiveness of introducing multimedia technologies in education activities and the teaching methods applications. The authors tested educational technologies on primary school children and their results show that in maths skills using programs or multimedia enhanced methods of teaching and that these methods can be effective in improve students' attention, especially if cartoons characters are used. Moreover, great importance is due to the effect of games and multimodal digital games particularly on children's achievements. In [Kynigos et al., 2010] the authors proposed full-body multimodal digital games set in an informal science education park. Their work, based on the framework of embodied learning theories, consists in different games on mathematical principles and tasks that required to students to explore and study, in a real practical word setting, rules and underlying concepts of mathematics, logic and geometry. The games they proposed were based on bodily shadow, weigh and position recognition and involved groups of players (from 3 to 12, depending on the game) interacting with digital representations on large projection surfaces. Results indicated that children perceived well rules and concepts on which the game were based and the longer they play the more their verbal interactions were driven to be concept-centred, instead of actions-centred.

2.3 Game design for education

One of the most challenging task in designing digital learning systems is that of engaging students, especially if children. An educational technology should not just be about providing information but also about facilitating students' experiences. Unfortunately, it seems that frequently educational technologies are used mainly as an information distribution channel that neglects the needs of students. To achieve and motivate students' new ways to develop learning technology should be found; the use of technology alone does not motivate, indeed, students that live in the middle of technology all their lives. To develop new learning experiences and new learning methods that engage learners, one approach is to introduce multimodal kinesthetic technology and serious games in education. *Play* is sometimes underestimated and evaluated as an activity essentially unimportant. However this view is mistaken. Play is one of the highest achievement of human species, with language, culture and technology. Furthermore, without playing, none of our achievements as humans would be possible. Researchers are increasingly recognizing the value of playing, within the educational field, even if the nature of "high quality" childhood education and gaming has been contrasted. One aspect that research highlights remarkably is that play is one fundamental ingredient in supporting a healthy intellectual and emotional social development of young children. The interesting on these studies is due to the emerging of a kind of consistent pattern. In all cultures, indeed, depending on available technology, there are manifestations of all the five types of play in which children engaged themselves: physical play, object-interaction play, symbolic play, pretence/sociodramatic play and games with rules. There are, however, variations from cultures and sub-cultures in children's attitude to play, arising from different cultural values and on childhood, play, genders and relationships constraints, social structures and so on. Gaskins and colleagues [Gaskins et al., 2007] have identified three general cultural perceptions of playing that seems to have an impact on children's patterns to play and their parents' involvement: a first situation in which game is tolerated but considered of limited value (e.g. in pre-industrial societies); a vision in which games are useful to keep children busy until they can be useful (e.g. pre-industrial societies either); and the last view, typical of European and American middle-class families in which play is considered as a fundamental part of child's work; it is encouraged and it is considered a relevant part of a correct child's development. Most of contemporary work on children's play are elaborated in the field of developmental psychology and based on the Vygotsky's theories, according to which playing is a crucial activity, because it makes crucial contributions to children's developing abilities, connected to the development of symbolic representations, such as language development, and in the development of cognitive and emotional processes to self-regulated themselves. Vygotsky's ideas are become increasingly recognized as research evidences show up that language and self-regulation are strictly connected [Vallotton and Ayoub, 2011] and they are, considered together, a powerful predictor of children's academic achievement and their well-being [Whitebread et al., 2011]. From the introduction of the International Classification of Functioning, Disability and Health (ICF) by the World Health Organization (WHO) in 2001 (but the last version is the 2016 version), play is identified as a major life area for children and youth under the "Activity and Participation" task. As we saw, it is widely accepted that play is important in child's development of physical, cognitive and social skills [Case-Smith and O'Brien, 2013]. Through play, children learn how to adapt to their environment. Lew and Campbell [Lew and Campbell, 2005] observed that every child in the world is engaged in musical play. Children play with music and also through music. Kemple and colleagues [Kemple et al., 2004] noted that children naturally sing, move and play. From a technological point of view, in the last twenty years, we have seen the rise of digital game and multimodal applications in entertainment, popular culture and as an academic field of study.

Computer games and technologies may create a new learning culture that corresponds better with students' habits and interests [Prensky, 2001]. Games are generally developed to satisfy the basic requirements of learning environments identified by [Norman, 1993] and can provide an engaging experience for students. The challenge for educational games is to engage and motivate players through direct experiences with the game world. They should provide possibilities for reflectively exploring phenomena and testing hypotheses, but, in recent years, they have been used in education primarily as tools for supporting information. Moreover, in action-based drill and practice games, players tend to try actions without reflecting on outcomes; in such games players may simply keep on trying and experimenting different actions until their scores improve. However, such "trial and error" behavior does not enhance learning. Researches as [Yildirim, 2017] emphasize the positive impact of game based programming education, as a strategy to motivate students and succeed by creating an environment in which they learn by doing. In this research the impact of digital game based programming course is stated with final course examination and questionnaires, resulting that students, aged 8-14 years old, have been positively helped by the game based program. Yildirim's research emphasizes the need for integration of educational theories, psychological and cognitive evidences, multimodal, multisensory and kinesthetic technologies and game design to be able to design meaningful and engaging educational games.

2.3.1 Game-based learning

Games are designed to generate a positive effect in players and they are most successful and engaging when they facilitate the flow experience. It is necessary to consider the actors that contribute to the flow: the gameplay is one of the other relevant aspects to pay attention to base an experiential multimodal game model. It is, in fact, the core of a game-based learning and its importance should not be underestimated. In [Rollings and Adams, 2003] the authors have defined gameplay as a series of actions that the players can take to meet challenges. Ambitions to design engaging educational games have probably often failed because educational aspects have displaced gameplay. The ability to solve problems, indeed, is one of the most important features of human skills [Holyoak, 1990]. Thus, one goal of education is to push students to experiment novel situations; in this framework problem solving can be considered as striving toward goal which is not immediately attainable. Games provide a meaningful environment to design problem solving tasks and experiences. Furthermore, problem solving can be associated with discovery learning. Learning environments like the games one, allow students to discover new rules and ideas rather than just memorizing the materials other people presented. An example of that is given by simulation games, a typology of game that offers the chance to the students to interact with the platform manipulating objects to make and test their hypotheses; as we can read in [Bruner, 1961], in this kind of game, while exploring the environment, the students become active learners and their motivation increases.

2.3.2 Experiential gaming model

Many are the advocates of the power that gaming educational can have to enhance the learning experience and results, but the real state of arts in educational game does not provide yet a full explored and understood methods to reach these goals, with game design often actually separated from learning. Mainly traditional educational models have been used to explain gaming model and few attempts have been made to integrate those fundamental aspects [Quinn, 1994, Kiili, 2005, Amory and Seagram, 2003]; moreover educational games often do not use the power of games as an interactive free media context where the player can actually try her hypotheses and improve her knowledge. Considering all these initial elements, in [Kiili, 2005] the author highlighted the need of a model that can be used to design and analyze educational games. He reported the model presented in [Amory and Seagram, 2003] as a first attempt of integrated framework of educational theory and game design, but without considering flow theory and gameplay aspects, presenting a too superficial model. The main purpose of the model presented in [Kiili, 2005]is to connect gameplay element to experiential learning, to facilitate the user experience. The author takes into consideration both constructivist [Phillips, 1995] and pragmatist [Kivinen and Ristela, 2003] theories on learning and propose a cyclic view of learning activities through direct experience done in the game world. In Kiili's framework a relevant role is played by the behavioral component of activity done during the learning process, focusing not just on the cognitive structures involved. The experiential gaming model proposed consists in an ideation loop, an experience loop and a challenge bank. The challenges based on the educational need are the core of the project: they should sustain the motivation and engagement of the player, proposing her appropriate challenges. To overcome the difficult task, the player generates solutions in the ideation loop part, working firstly on a proper creativity moment, and then the idea generation. The first phase, the creative one, according to [Finke et al., 1992] is more successful undertaken without considering constraints of the system, leading in that way to novel solutions. After the ideation phase, the game should be playable, provide clear goals and feedbacks to user and have a clear interaction and flow design. From a learning point of view, it is important that the players can test different kind of solutions to expand their knowledge on the relevant subject. Designing the challenges in the ideation loop it is important to consider the tempo of the tasks, that has to be balanced with the player's relevant skills. The point of the games is to keep the player in a flow state increasing the skills level of the game while the player's skills increase, considering also that it is impossible to predict player's velocity to master new knowledge, an element that makes particularly difficult the designing of the game. An interesting work was done by [Ketamo, 2003], who developed an adaptive geometry game that focused on player's behaviors to provide the appropriate challenge to the player. The positive result was particularly relevant for those players who had a low geometry level skill and that benefited of the adaptive component of that solution. Relevant for the educational context, are the balance and the cognitive load of the game. The main goal of a balancing game is to provide a consistent game, allowing players to exploit different solutions to take advantages from them. Moreover, the idea of a positive feedback loop (e.g. the action of a player ahead in the game is in some way easier:

the progress in the game is rewarded positively), is controversial with the goal of education and it tends to create a situation in which player with higher skills (and score) get higher, instead lower-level players tend to lower results each time. Although good performance should be rewarded, also the progress of lower skills players must be somehow supported. One solution is to adapt the game at the player's skills, as already suggested before. In addition, a well-balanced game is a good answer to ensure that all the players can catch up the right challenges and the other players' levels. A relevant problem of multimodal and multimedia learning materials is the limited working memory capacity, often overloaded due to inappropriate design, interaction or presentation [Kiili, 2005]. [Mayer and Moreno, 2002] presented a cognitive theory of multimedia learning materials that considers the working memory limits for both visual and auditory processing., stressing that the working memory limits the amount of information processable in each channel at the same time [Sweller, 1994]. Integrating haptic feedback in learning experience is a way to reduce extraneous cognitive load, because technology that uses tactile sensory information can simulate kinesthetic sensations so that the players can feel and not just look at the game [McGee and Cohen, 2003]. Haptic and movement support an embodied and immersive experience, in a way that let the players to learn in a more realistically way, new materials. The general assumption of multimodal and development is that the ability to increase working memory capacity can be enhanced by using multisensory information processing channels simultaneously. As pointed out by [Kiili, 2005], in 1998 Sweller and colleagues [Sweller, 1988] identified three different sources of cognitive load. The first one is the intrinsic nature of the material, identified as intrinsic cognitive load, the second one is the way in which the material is presented, that is the extraneous cognitive load, and the last one is the effort needed for the construction of the schemata, and this is named germane cognitive load. The intrinsic cognitive load refers to the nature of the task or the subject to be learnt: if there are numerous elements related to one another, the cognitive load will be high. According to cognitive load theory, instructional design cannot affect this type of cognitive load. The most important aspect of cognitive load theory for educational game design is the extraneous cognitive load, that is an unnecessary cognitive load and it is directly determined by instructional design. If the game is not well designed, this type of load will be high because learners will be involved in numerous and not relevant cognitive processes. However, it is not guaranteed that a reduction of extraneous cognitive load let to a deeper knowledge process. Unused working memory capacity should be used to optimize the third type of cognitive load, the germane one, by stimulating the player to process the problems provided deeply. In conclusion, cognitive load should be optimized in games, deleting all the irrelevant elements, using multisensory information processing framework, providing user interfaces and challenges that actively support learning.

2.3.3 Examples of multimodal game-based applications

The consensus on the positive effect that a well design educational gaming can have, can be widely investigating in literature. In 2012 Shin and colleagues [Shin et al., 2012] conducted a double study to investigate the positive impact of game technology on learning. One of the problem in educational technology research is the lack of dataset to validate the contribution of technology to students' learning, because too often the collected data are considered too fragmented ad unsystematic [Condie, 2007]. In the [Shin et al., 2012] study, the authors investigated the positive effects of game technology on student learning in mathematics, using two different datasets collected from slightly distant subjects, all aged 7-8 years old. The authors applied different types of statistical analysis on the dataset and their results revealed that using a technology-mediated game was beneficial for all the students involved, no matter of different ability levels or personal differences. Game technology have supported consistently positive results particularly concerning motivation, persistence, curiosity, attention and attitude to learning. In another work of 2015, Tsai and colleagues [Tsai et al., 2015], investigate the use of Kinect sensor as interactive technology that can assist and improve teaching and learning. Particularly, the authors tested the efficacy of Kinect game-based learning of mathematics concepts. The authors used the ARCS model [Keller, 1999] to provide a kinesthetic pedagogical experience of spatial skills, motivating students and enhancing effectiveness. In Keller's method of motivational design, there are identified four steps for learners to be and remain motivated in the learning process: attention, relevance, confidence and satisfaction. The first two steps are the core of the model. In accordance with the theory, attention refers to the interest the learners show on concepts and ideas taught. For the relevance, it is important to use a language and provide examples that learners are familiar with. Confidence focuses on establishing positive expectation for players for achieving success. For the satisfaction, it is important to provide to the learners the right reward for the learning process. This is why feedback and reinforcement are important elements for learners to keep the motivation high. Designers should work to enhance motivation and germane cognitive load to enhance learning effectiveness [Woo, 2014]. Following Keller's model, the authors implemented a game centred on four components: interactivity (to elicit attention), representation and storytelling (for the relevance of the context), rules and goals (for players' confidence), and consequences (players' feedback and satisfaction).

The data collected by the authors indicate that usability plays a vital role in the initial adopting of a technology and learners believed that game technology can facilitate learning. The learners also reported that interactive system teaching helped them to enhance their motivation and was helpful to learn spatial knowledge. Digital and multimodal technologies have been used in last years in various forms of training, teaching and learning, even in medical education [Rosenberg et al., 2010]. Two experiments conducted by Erhel and Jamet [Erhel and Jamet, 2013] showed that a serious-game environment can enhance learning and motivation, and thus learning performance, providing features that let learners to actively process educational content. Moreover, in the rehabilitation field, in [Chuang and Kuo, 2016] the authors present a game-based therapy for physical training of children with Sensory Processing Disorder (SPD) to enhance kinesthetic intelligence and helping them facing learning challenges. Sensory Processing Disorder is a neurological condition in which the brain cannot accurately process sensory information, and thus patients can have behavioural and learning disorders. The actual identification of this disorder can be categorized in four patterns: "Visual perception and Auditory-Language Disorders", "Tactile Defensiveness", "Disorders involving the Vestibular System" and "Developmental Dyspraxia" and children affected by SPD usually have several symptoms of these four patterns [Ayres and Robbins, 2005]. [Chuang and Kuo, 2016] present in their work a serious-game to help children with SPD to be more active, improving sensory integration and processing and linking this improvement to academic learning and social learning needs. The study focused on aerobic and balance games. The results show that the proposed game-model has better effects in improving the kinesthetic learning of SPD children compared to traditional therapy, especially for balance and jumping. The proposed system can provide a friendly environment for a life-long training, improving not only physical coordination, but also learning and social skills. The subjects show improvements also in manipulating objects abilities in daily life and better walking.

2.3.4 Game design for mathematics learning

Educational gaming for mathematics and sciences is not as simple as just let students play some game and expecting a substantial changing in their achievements and goals. Educational games need to be designed with attention to contemporary pedagogy, including adaptivity to task difficulty to learners' capabilities, metacognitive reflections and consideration on the interaction among learner, game environment and the classroom [Young et al., 2012]. Moreover, considerations on serious games often consider only standard interface, reducing the affordance for multisensory learning. In [Devlin, 2013] is proposed an interesting survey on online video games and apps that claim to help learning mathematics concepts. For the authors there are few precise parameters that can be used to evaluate a learning game. A good mathematics learning game or app should avoid:

- To confuse mathematics itself with its representation (e.g. symbols);
- Presenting mathematical activities in a separate way from game mechanics itself;
- Relegating mathematics to a second aspect of the game;
- Pay attention not to communicate that mathematics is an obstacle in the way to do more interesting activities;
- Encouraging students to try answers too quickly or without thinking enough;

Most of the games, in the author prospective, focus not enough on learning and understanding, but on mastering basic skills, such as "multiplication tables" or fractions. This kind of game does not actually provide learning, but it does a good use of video game technology to take out of the classroom specific knowledge acquisition. Following the American "Adding it

Up: Helping Children Learn Mathematics", published by the National Academies Press in 2001, [Council et al., 2001], there are, indeed, five elements that contribute to the so-defined mathematical proficiency in children education. These elements are conceptual understanding, that refers to the comprehension of mathematical concepts, operations and relations; procedural fluency, defined as a skill on do mathematical procedures efficiently, with flexibility and appropriately. Strategic competence defined as the ability to represent and solve mathematical problems in real situations. Adaptive reasoning, which refers to the capacity for logical thinking, reflection and explanation. Productive disposition, the ability to see mathematics as useful and worthwhile, combined with confidence in one's ability to master the material. To build a successful mathematical games, according to Devlin, requires a comprehension of what mathematics is, how and why people learn and use mathematics in everyday situations, how to get and keep them engaged in a learning strategies, and how to represent mathematics concepts in all the platforms in which the game is played. Just as music is created and enjoyed with mind, so mathematics does. Both mathematics and music are mental activities and in both music and mathematics symbols are representations of dynamic mental processes. As music instruments provide interfaces that are native to music and hence easily a natural use of them, a properly designed mathematical game can provide interfaces to mathematics concepts native for those concepts, so easy and natural to use. Learning by doing is certainly an important aspect of music education and practice that can be transposed to mathematics educational game. Devlin also highlights that, likewise the music instrument learning, it is not true that for learning mathematics it has to master basics skills before progress, but rather it's a matter of how the student acquires the mastery that makes the difference. As when a music student can go back and forward to master easily techniques that she needs, so the same mechanics can be applied to mathematics learning. The author gives some examples of games that follow these general principles: MotionMath and MotionMathZoom, for example, which use motion sensors of smartphone or tablets to directly interact with numbers. DragonBox that focuses on learning algebra in a puzzle where a dragon box has to be isolated. KickBox uses physical concepts – as position of objects in space, to teach some maths concept. More recent results show how music also integrated in virtual reality can improve mathematics learning. In [Lim et al., 2018], the authors propose a study to investigate teachers' experiences and perceptions of using virtual reality game for elementary maths education. The game they proposed, integrates a beat-making into a maths learning game for fractions. The results they showed, highlights that the concept of fractions is effectively represented via beat-making in an immersive game. The study also show that musical term clarification and adaptive, haptic manipulation of the game are salient features that influence game-based learning.

2.3.5 Game design for visually impaired children

There are in literature several studies using computer applications to improve learning in children with visual disabilities [Winberg and Hellström, 2000], [Baldis, 2001, Sánchez and Lumbreras, 2000, Sánchez et al., 2001, Sánchez et al., 2004, Lumbreras and Sánchez, 1998, McCrindle and Symons, 2000, Mereu and Kazman, 1997]. Most of the studies are based on interactive software not able to completely adapt to children needs. In [Lumbreras and Sánchez, 1998], indeed, the authors propose to use audio feedback to enhance spatial cognitive structures and found that spatialized sound can be used to develop spatial navigation skills in a multimodal virtual environment. In other studies, such as [Mereu and Kazman, 1997], it was proposed a designed experience of audio stimuli to simulate visual cues for blind learners. Using 3D audio interfaces, indeed, help blind people to localize specific point in 3D space. The precision of blind people in that circumstances were high, but they are slower than sighted people, concluding that navigating virtual environments with only sound can be more precise for blind people compared to sighted ones. In [Lahav and Mioduser, 2001] the authors used a sensory virtual environment through force feedback joysticks simulating real life environments, as school or work places. They proved the hypothesis that providing an appropriate spatial information through compensatory channels can improve performances of blind people. In [McCrindle and Symons, 2000], researchers used force feedback joysticks as input device of a Space Invader game replicated with 3D sound. The authors tested blind and sighted children do the same experience, concluding that traditional computer game can be played by blind children using 3D audio system. In [Sánchez et al., 2004], the authors proposed a study on design, development and usability of an interactive virtual environment based on audio to enhance short-term memory and assist mathematics learning of children with visual impairments. Blind children participated in the design and usability tests during and after implementation. The authors implemented a principal interface for children with residual vision and designed the game to be also only-audio to be used autonomously by completely blind children. The interaction with the interface elements as buttons and text screen was possible through keyboard. Each interaction with the game triggered an audio feedback and a high visual contrast screen for children with residual vision. The child must move through a grid and for each cell it was associated a music toned to identify the position in the grid. Sound is listened during the movement through cells. When a token was opened the associated element was presented by triggering the audio feedback. For example, if the image was a car, a real traffic sound was trigged. Opening the correct pair of token, a feedback was provided. Finally, when all pairs are made, time, score and feedback were given. Alternatively, to the use of a keyboard for the interaction, the authors designed the game to be playable also with joysticks or on tablets. Audio Math was designed to enhance memory in visually impaired children, through practising audio and visual – if possible, memory. The task implied to exercise audio/oral, visual/oral, audio/graphic, visual/graphic memory. The game also emphasizes learning concepts such as equivalences, correspondence relationships, differentiating tempo-spatial notions. Results showed that sound can be a powerful interface to develop and enhance memory and mathematics learning in visually impaired children. Sánchez and Elías [Sánchez and Elías, 2009] proposed an interactive audio-based multimedia software for children with visually impairments to support the learning of science. The authors focused on how to learn science using audio as primary input and how to develop challenging and encouraging applications that, at the same time, can assist learning. The authors proposed a customizes navigation model for end-users and built a generic software for role-playing games. They evaluated the usability of the software and the cognitive impact of it, verifying that the software stimulates an independent and spontaneous use of it by users. The application was overall considered appealing, challenging and encouraging for learning science by visually impaired children. Lumbreras and Sánchez [Lumbreras and Sánchez, 1998] worked on interactive 3D audio hyperstories for blind children several years ago. They realized AudioDoom, an interactive model-based software for visually impaired children. The protype was field-tested with several blind children. The results found indicate that spatial sound experiences can create spatial navigable structured in visually impaired children's mind. In 2001, Sánchez and colleagues [Sánchez et al., 2001] presented they first results on audio interaction for blind children. Blind children can build mental structures only based on audio stimuli. Spatial imagery is not purely visual and can be constructed and transferred through spatialized sound.

Chapter 3

A multisensory computational model: an adaptive multimodal system for non-verbal and affective behavior

3.1 Design based research approaches in educational technology

Design-based research (DBR) is a generally well-known research design approach that suits well multisensory educational technology. It was also of relevance for the educational research contexts, as it aimed to impact practice, offering design interventions through the design process [O'Leary, 2014]. In this approach "researchers manage research processes in collaboration with participants, design and implement interventions systematically to refine and improve initial designs, and ultimately seek to advance both pragmatic and theoretical aims affecting practice" [Wang and Hannafin, 2005]. To ensure veracity and generalisability of the findings, the research should be grounded on relevant literature and real world contexts. In both the case studies that will be presented in next sections, this kind of approach has been used. Furthermore, research in External Cognition shows how different properties of external representations (e.g. audio) influence our cognitive way of interaction and, consequently, of learning. Finally, the other relevant context was multimodal perspective, meaning how people use multiple semiotic resources for communication and meaning making. Real-world setting was fundamental for the good outcome of both the projects that will be presented in the second part of the thesis, since it let us to address needs that are evident in nowadays educational practice, in both the contexts considered. Observations of everyday class activities, together with teachers interviews were central to this approach. In terms of credibility of research, the participatory design provides confirmability, through iteratively engaging with teachers, documenting findings from workshops, interviews, observations and questionnaires, using these strategies to move forward in the research. While DBS prescribes strategies for respecting natural contexts of activity within the inquiry, there are no particular prescriptions for methods, and typically it is used a mixed methodology process, integrating both qualitative and quantitative components. The mixed methods ensure both *objectivity*. using also multiple kinds of data [Akilli, 2008], and *authenticity*, describing, for example, events in a manner that it sounds correct to the stakeholders' experience. In participatory design, ethnographic investigation and fieldwork-inspired techniques are increasingly used. Without observing practice *in situ*, a common risk is to find the perfect solution to the wrong problem [Kensing and Blomberg, 1998]. Observations enable designers to gather information about who the users are, what kind of task they perform and the contextual setting [Blomberg et al., 1993]. Participatory design let to understand knowledge by doing. The user-centered design process focused on the thing being designed (e.g. the object or interface) looking for ways to ensure that it meets the needs of users [Sanders, 2002]. Participation is a key-element in social interaction, such as classroom activity, and is essential to consider it in good design practice. Three basic strategies are present in almost all participatory design research [Spinuzzi, 2005]: a) initial exploration of work; b) discovery processes; c) prototyping. The advantage of such methodology is that techniques as mock-ups, simulations of relation between work and technology, workshops, design games, case-based and cooperative prototyping, help people taking part to unlock tacit knowledge that informs successful design to interaction [Kensing and Blomberg, 1998]. In elaborating practical problems of work through a continuous process of analysis and design, participants 'work up' alternative futures [Crabtree, 1998]. In particular, children may find it difficult to communicate to adults exactly what they are imagining and prototyping offers a concrete way to discuss ideas [Druin, 1999].

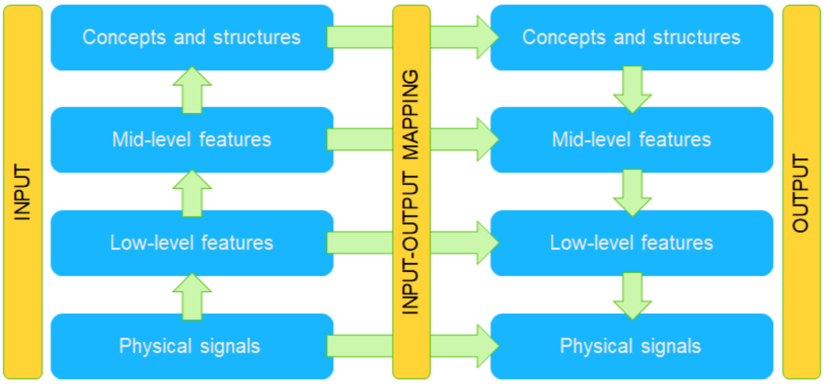
3.2 Conceptual computational Framework

Considering full-body technology applied to educational contexts, we firstly need a computational framework suitable to be applied both to these particular fields and to movement and gesture technologies. For the analysis of body movement expressions, we based our work on the *multilayered conceptual framework* developed by Camurri and colleagues (Fig. 3.1), initially proposed in [Camurri et al., 2003], [Camurri et al., 2005] and recently extended in [Piana et al., 2013]. In both the case studies on which the author actively work, this framework was considered as the computational basis of the approaches proposed. In other words, the author did not contribute, actively, neither to the definition or implementation of the model, but applied it in two concrete case studies scenarios. In the next pages, the model will be presented to make the reader familiar with it, however, specifying once again that it is not the result of the author's theoretical or implementation work. The framework supports multimodal systems and its ultimate goal is to build a set of high-level features that could help to describe and to understand the non-verbal component, e.g., the emotional state, expressed by the users and to provide them a feedback based on their performance. The framework defines four different layers of body movement features (from Layer 1 to 4):

- Layer 1 - physical signals (e.g., pure positional captured sensors or motion capture data and simple kinematics, muscle activity)
- Layer 2 - low-level features (e.g., energy)
- Layer 3 - mid-level features, maps and shapes (e.g., impulsivity)
- Layer 4 - concepts and structures (e.g., emotion, internal states)

Starting from the recorded data (the physical levels), low-level features are initially computed. Their analysis then evolves into mid-level features with the aim of identifying a high-level features (i.e., positioned at the third or fourth layer of the framework) that better characterises the relation between the body expression and learning factors, e.g., self-efficacy, amount of engagement, attention or stress level of a child while learning. Considering the analysis of expressive gesture in human movement, significant examples are (i) the amount of movement detected by the video-camera (Motion Index, formerly Quantity of Motion), computed as the normalised area (i.e., number of pixels) of a Silhouette Motion Image (SMI), i.e., an image carrying information about the variations of the silhouette shape and position in the last few frames, and (ii) the amount of contraction/expansion (Contraction Index), computed as a ratio of the area of the bounding rectangle and the area of the silhouette. Important features are those inspired by the Effort dimensions described in Rudolf Laban's Theory of Effort [von Laban, 1975], [Laban and Lawrence, 1947], e.g., directness (computed as the ratio of the length of the straight line joining the first and last point of the trajectory followed by a moving subject and the length of the trajectory itself), impulsivity, and fluidity (both computed from the shape of the velocity profiles). In the case of music, low-level features may also be related to tempo, loudness, pitch, articulation, spectral shape, periodicity, dynamics, roughness, tonal tension, and so on: a similar conceptual framework and a taxonomy of audio features can be found in [Leman et al., 2003], [Lesaffre et al., 2003]. There can be, indeed, found analogies among features in movement and in music, e.g., amount of motion – loudness, contraction/expansion – melodic contour or spectral width, bounded, hesitant movement – roughness. In the mid-layer, data from several physical and virtual sensors are therefore likely to be integrated in order to perform such a step. Each gesture is characterised by the measures of the different features extracted in the previous step (e.g., speed, impulsiveness, directness, etc. for movement, loudness, roughness, tempo, etc. for music). One problem that at this step of the analysis has to be solved is to choose the right segmentation window, since in human expression finding right boundaries is not always evident, in automated analysis several variables have to be taken into account, such as different temporal scales (expressivity may needed longer time to emerge) and complex body joint coordination/entrainment. The last layer, Layer 4 (Concepts and structures), is devoted to extract high-level expressive content from expressive gestures. It can be organised as a conceptual network mapping the extracted

Figure 3.1: Computational conceptual framework



features and gestures into (verbal) conceptual structures. One example can be to focus on emotional labels (e.g., the basic emotions anger, fear, grief, and joy) or on dimensional approaches, such the well-known valence-arousal space, or other dimensional models [Roseman, 2013], or those especially developed for analysis and synthesis of expressive music performance, e.g., [Juslin, 2000], [Vines et al., 2005]. Machine learning techniques can be used at Layer 4, including statistical techniques like multiple regression and generalized linear techniques, fuzzy logic or probabilistic reasoning systems such as Bayesian networks, various kinds of neural networks (e.g., classical back-propagation networks, Kohonen networks), support vector machines, decision trees. The conceptual architecture sketched above is conceived for analysis (so the direction is upward). Anyway, a similar structure can be employed also for synthesis (downward). In our case studies, to whom the next section is dedicated, we did not actually use Level 4 of the model, since we did not work on emotional states recognition. Nevertheless, especially in the weDRAW project 4, we will make extensive use of low and mid level features for the detection of children cognitive and affective states during their interaction with the technologies proposed by the project.

Whereas the conceptual framework was initially conceived for dance performances, the specific challenge we faced in both the projects, described in the next section, consisted in extending it to learning settings, both for primary school mathematics learning and adult students music education (with a particular focus on violin playing). Moreover, in Chapter 6 some works are presented in which this multi-layer motion model was used to identify body movement features helping in detecting possible dangerous movements and postures that could lead in the long-term to injuries and right bowing techniques. We exploited the movement libraries implemented in EyesWeb to extract low-level features (e.g., velocity, acceleration, number of pauses, kinetic energy and so on) measured on the most important body joints. Low level features were then exploited and used to model and estimate features at the highest levels, such as Confidence and Reflective Thinking in learning tasks.

3.2.1 Capturing devices

Concerning capturing devices (Layer 1 in the conceptual framework), in both the case studies we used a high-end high-performance motion capture platform (Qualysis), a low-cost solution (Microsoft Kinect), and IMUs and physiological sensors (e.g., Myo and Notch sensors). Parts of the body that are deemed particularly relevant for analysis include head, trunk, right arm, left arm, right hand and left hand (to the extent at which this is feasible). In the TELMI project, concerning music learning, an analysis of the whole body was also useful, namely for detecting of postural components in violin performance.

3.2.2 Body movement features in learning contexts

The features used in both the projects, were selected according to their use in previous studies, for their application in analysis of children non-verbal communication and for the analysis of music performance with other musical instruments than violin.

3.2.2.1 Body movement features in music learning

Following the second case study, the one on music learning (see Chapter 6), the analysis of body movement in teaching and learning violin playing was receiving interest from the scientific community since long time (e.g., see [Jacobs, 1969] for an early study). Whilst other studies mainly adopted qualitative approaches based on observation of video recordings (e.g, see [Ruggieri and Katsnelson, 1996]), nowadays, with the increased availability and accuracy of motion analysis technologies, a greater attention is giving to quantitative approaches. For example, [Goebl et al., 2014] summarized recent quantitative measurement and analysis techniques of three domains of musical expressiveness: body motion, musical sound, and listeners' continuous response to musical sound. In particular, they outline computational methods to quantitatively assess expressive aspects of the body movements of the performing musicians. Concerning violin playing in particular, research often addressed kinematic features of body movement and specific aspects of the motor behavior. For example, Turner-Stokes and Reid [Turner-Stokes and Reid, 1999] investigated rotations of upper limb movement in the bowing arm. Visentin and colleagues [Visentin et al., 2015] analysed biomechanics of left-hand position changes. The i-Maestro project, which was introduced in Chapter 2, explored the relationships between the bow, the instrument and the body [Ng and Nesi, 2008]. A rather under-investigated area is the analysis of expressive features in violinist body movement. Glowinski and colleagues [Glowinski et al., 2014] studied how a violinist's body expression can adapt to a variety of expressive styles. Using a motion capture system, they captured body segments and the instrumental dynamics (head, arms, torso, sacrum, and violin) of three violinists playing in metronomic, emphatic, and concert-like styles, and they applied a General Linear Mixed Model to evaluate how the expressive style impacts body segment displacements. With respect to inter-limb coordination, Baader and colleagues [Baader et al., 2005] measured quantitatively bimanual coordination in violinists playing repeatedly a simple tone sequence. In a follow-up study, Kazennikov and Wiesendanger, [Kazennikov and Wiesendanger, 2009], used motion capture data to investigate basic bimanual movements played by violin amateurs and professionals to check whether position and string changes may influence bowing (right)-fingering (left) coordination. Visi and colleagues, [Visi et al., 2015], empirically explored the shared knowledge of the gestural repertoire of violin performance by people having no previous experience in playing that particular instrument. Violin was chosen since it is a well-known musical instrument. They computed Mocapgrams [Jensenius et al., 2009] and Periodic Quantity of Motion (PQoM) to analyse periodicity and phrasing from motion capture data. PQoM [Visi et al., 2014] is a motion descriptor, inspired by Quantity of Motion [Camurri et al., 2003] useful to observe how movement relates to rhythmic aspects of the music. Some recent studies also addressed higher-level movement qualities in violin performance. A useful source of inspiration is research on expressive movement qualities focusing on other musical instruments. Just to give a few examples, Broughton and Stevens [Broughton and Stevens, 2012] investigated whether the basic components of Laban analysis are reflected in perceptual judgments of recorded performances of marimba players. Davidson [Davidson, 2012] studied the expressive components of bodily movement in both solo and ensemble musical performance, focusing on flute and clarinet performers and on a pianist. Demos and colleagues [Demos et al., 2014] applied Recurrence Quantification Analysis (RQA) to examine performers' movements (the postural sway of two trombonists) and relate them to the musical structure and to the musician's expressive intentions. They also showed how Detrended Fluctuation Analysis (DFA) can be used to identify synergies and discover how they are affected by the performer's expressive intentions. Most of the above-mentioned work are grounded on motion capture data. Approaches, however, exist based on RGB-D devices or other sensor systems. For example, Hadjakos and colleagues, [Hadjakos et al., 2013] presented a method for motion analysis of musical ensembles based on head tracking with a Kinect camera. In particular, they applied their approach to a violin duo performance. Hsu and colleagues [Hsu et al., 2014] developed a multi-channel system to record the audio and the electromyography (EMG) signal simultaneously during a violin performance. They also developed algorithms to analyze the music performance and discover its relation to player's motor sequences.

3.2.2.2 Body movement features for children learning task

Bodily gestures exhibited during learning experience can also be thought of as physical bodies (becoming study materials for learning the enacted concepts) to abstract concepts [Kim, 2011]. Given the importance of gestures and body movements, therefore, we discussed their use as modality for automatically detecting learning states in the weDRAW project [Olugbade et al., 2017]. As we stated in the cited work, although there have been several studies that have investigated automatic monitoring of cognitive and affective states in the context of learning (e.g. [Kapoor et al., 2007], [Bosch et al., 2016], [Rodrigo and Baker, 2009], [McQuiggan et al., 2008], [J. Grafsgaard and Lester., 2015]), a gap still exists in understanding which of these states is of benefit to monitor. We decided to focus our attention on three stated from the literature: self-efficacy, curiosity, and reflectivity/impulsivity. We focused on these states as they had not received, in prior research in affective computing, much attention compared to other learning states such as frustration, concentration, and boredom. For example, beyond studies such as [McQuiggan et al., 2008], [J. Grafsgaard and Lester., 2015], there has been limited discussion on learning facilitation opportunities that could be created in monitoring self-efficacy in digital learning systems. In learning, self-efficacy, curiosity, and reflectivity are epistemic states, i.e. related to cognitive appraisal or processing of information being learnt, different from achievement states (activity or outcome related), topic related states, and incidental states which emerge from experiences outside the classroom [Pekrun and Linnenbrink-Garcia, 2014]. Thus, they are central to learning and from an interaction design point of view, it is critical to understand how they may be promoted, moderated, or supported in digital learning environment. The importance of self-efficacy in learning and its influence on the amount of effort that a learner will put into the completion of the learning task in the face of barriers [Bandura, 1997]. In [Schunk, 1981], for example, the authors showed that children with higher level self-efficacy for arithmetic problem solving spent a significantly greater amount of time attempting the problem than those with lower levels. Self-efficacy also has an influence on curiosity to learn as was found in a study by [Litman et al., 2005] with undergraduate students in a trivia task. The authors found that in cases where participants indicated not knowing the answer or that it was on the tip of their tongue, the reported level of confidence in the ability to identify the correct answer if given multiple options predicted the level of curiosity for the correct answer. The cognitive strategy used in problem solving, reflectivity versus impulsivity, also affects learning outcomes. In contrast to impulsivity, reflectivity enables learning and problem solving as it stimulates focus of attention, more analytical cognition, and use of helpful problem-solving strategies [Rozencwajg and Corroyer, 2005]. Indeed, in a study by [Navarro et al., 1999], children who were impulsive performed significantly worse in arithmetic problem-solving tasks than their reflective peers. Previous studies provided evidence of the feasibility of automatic detection of levels of self-efficacy and curiosity. For example, [McQuiggan et al., 2008] used physiological signals and interaction data (such as time spent and progression towards the goal) for automatic detection of levels of self-efficacy. They obtained accuracy of 0.87, 0.83, 0.79, and 0.75 for detection of 2, 3, 4, and 5 levels respectively. Similarly, in [J. Grafsgaard and Lester., 2015], facial cues were used to automatically detect levels of self-efficacy for middle school and college students respectively. In [Arroyo et al., 2009], a combination of facial cues, skin conductance, computer mouse handling, sitting posture cues, and interaction data (e.g. progression towards goal) were used to automatically classify 5 levels of self-efficacy. While these studies provide evidence of the possibility of detecting these states using physiological sensors, in weDRAW project we focused on bodily gestures and body movements, conceiving those modalities as primary components of traditional learning environments and may be a window into the mind of a learner [Goldin-Meadow, 1999]. Further, unlike the traditional affective computing modalities (face and voice), bodily gesture and body movement encapsulate information about the action tendency of the learner towards coping with or addressing the experienced state and so offer unique insight into subjective experiences [De Gelder, 2009]. Relationships have indeed been found between bodily gesture and body movement behaviour and self-efficacy and reflectivity-impulsivity. For example, in [Olugbade et al., 2018a], it was found that movement performance cues enabled observer assessment and automatic detection of levels of movement-related self-efficacy. Speed of movement, range of motion, muscle tension, asymmetry in movement, and movement fluidity were particularly found to be useful cues. Although their work largely focused on a clinical population, some of these cues are similar to those found in [Keogh et al., 1981] with child athletes performing gymnastic routines. Exploratory behaviour may, thus, be a possibly cue of this state. Movement performance has also been found to be related to reflectivity in [Keller and Ripoll, 2001] where reflective children were found to be better at motor tasks (using a racquet to hit a ball towards a target) than impulsive children.

3.2.3 Movement features computing

Below, we provide a list (with brief descriptions) of features for which computational algorithms have been implemented through the years and used in the case studies presented in next section. The associated algorithms are integrated in the EyesWeb XMI platform. The platform is developed in software modules (EyesWeb blocks), that can be easily assembled (e.g. connecting blocks together with a "drag and drop" logic) to create programs, called patches that are implementation, in this context, of multimodal recorders, research analysis tools and serious games. In most of the functions of the EyesWeb XMI platform, motion features are computed from three-dimensional (3D) user tracking information captured by optical motion capture systems (other motion capture systems are also considered) and are mainly computed from the user's body posture and on the trajectories drawn by the user's joints. In particular, user's tracking is mainly based on depth sensors (e.g. Microsoft Kinect) and webcams. Trajectories can be trajectories coming as result of the process of tracking something of interests in the real world (e.g., the trajectory of a single joint for example of the head of a user) or trajectories in virtual, semantic spaces where the space regions are representing expressive qualities. Real world trajectories can be defined as collections of consecutive positions that represent the movement performed. In detail, a movement trajectory is a sequence of 3D points (p1, p2, ..., pN), where a single point pk is a three-dimensional point identified by 3 coordinates in space i.e., pk = (xk, yk, zk). In particular, for both the case studies we were interested the 3D positions of:

- the barycentre of the body, for a movement spatial analysis.
- relevant body segments/joints, with particular reference to head, trunk, right arm, left arm, right hand and left hand, for movement expressivity analysis.
- selected body angles (e.g. arm-shoulder-trunk angle) to monitor particular musicians' body posture.

Features can be computed on the entire length of the trajectory or on the last M points of the total N, using sliding windows (to compute the time evolution of a particular feature over the entire trajectory). In the case of wearable devices (IMU sensors), the features are mainly computed from 3D angular information. The body movement features extractions are listed in Tables 3.1-3.2, organized according to the earlier discussed hierarchical framework. The list contains function for both the EyesWeb XMI platform and the stand-alone library (the latter are indicated by a *). Starting from low-level features (Layer 2), they describe movement at a local time-scale (i.e., they are instantaneous qualities). They consist of time-series computed either on each raw data sample or on a small sliding window. They capture both postural and kinematics aspects. We distinguish between features computed from motion capture data and from video analysis. Some features can be computed from both motion capture and video (of course with different accuracy). As we stated for the computational theoretical framework, also for the features here presented the author wants to clarify that these features are resulting of years of work of all the people that developed EyesWeb XMI Platform. For the case studies of weDRAW and TELMI, the author has limited herself to compute and analyse the features described in the way here presented and has not implemented any new ones.

Kinematic features from motion capture	
Feature Name	**Description**
Position	The 3D position of relevant body segments/joints, with particular reference to head, trunk, right arm, left arm, right hand and left hand (to the extent at which this is feasible).
Rotation	Rotation of relevant body parts, with particular reference to wrist. Rotation of a single body segment with respect to other ones (e.g., arm with respect to hand, shoulder with respect to arm, and so on) are also considered.
Velocity	Velocity of the aforementioned body segments/joints.
Acceleration	Acceleration of the aforementioned body segments/joints.
Jerk	Jerk of the aforementioned body segments/joints. Jerk is directly connected with biomechanical efficiency [Flash and Hogan, 1985] and is involved in the computation of Smoothness [Hogan and Sternad, 2007].

Direction	This mainly refer to head and gaze direction. Head and gaze direction already proved significant in analysis of string quartets and string orchestra sections (e.g., [Glowinski et al., 2013], [Gnecco et al., 2014]).
Curvature	Local curvature of the trajectory of the aforementioned body segments/joints. Curvature may be exploited as an indicator of movement directness and it can be employed in gesture segmentation algorithms.
Smoothness	Smoothness of the movement of the aforementioned body segments/joint: this is computed according to biomechanical models [Hogan and Sternad, 2007] and is propaedeutic to computation of Fluidity (Layer 3). Smoothness and Fluidity are indicators of the continuity of a player's movements.
Kinetic Energy	The kinetic energy of the cloud of 3D tracked point. This is computed by associating a mass to each point, according to anthropometric tables (e.g., see [Dempster and Gaughran, 1967]). Since the relevance of right hand in moving the bow, it could be worth to compute its kinetic energy separately and to compare it with the overall kinetic energy.
Kinematic features from video (RGB-D sensor or camera	
Feature Name	**Description**
Position	The 3D position of relevant body segments/joints, with particular reference to head, trunk, right arm, left arm, right hand and left hand, as long as they can be tracked robustly.
Velocity	Velocity of the body segments/joints that can be tracked reliably.Acceleration and jerk are usually noisy, unless accelerometers are used to capture them directly.
Direction	Head movement direction can be computed if head tracking can be performed reliably. Overall direction of body movement can also be computed (e.g., by means of optical flow).
Curvature	Local curvature of the trajectory of the body segments/joints that can be tracked reliably.
Motion Index	Motion Index (a.k.a Quantity of Motion) provides an approximation of Kinetic Energy in terms of amount of movement the video camera detects [Camurri et al., 2003].
Length and duration	Trajectory length in term of recorded samples, and duration in seconds.

Average velocity	Average velocity of a single or group of body joints or body barycenter
Average acceleration	Average acceleration of a single or group of body or body barycenter
Movement main direction	A directional conveyor, calculated by regression, of the main direction followed by the trajectory.
Pause detection	Detection of movement of pause, useful for segmentation of trajectories and movements
Direction change	Detection of a change in direction i.e., a measure of how, the trajectory of movement detaches abruptly from a mean value.

Table 3.1: Low level features

Name	**Description**
Directness	Directness is a measure of how much a given trajectory is direct or flexible. It can be used to infer how much the children certain or uncertain the movement who did. We will use Directness and curvature as component of mid-level features (Fragility, Rigidity) and high-level features (Stress, Uncertainty).
Openness	A measure of the extent at which body segments/joints are close or far from the barycenter. It is computed from the distance from the barycenter of each 3D tracked point.
Tension	A measure of the amount of torsion in a posture. Tension is computed from the angles between the adjacent lines identifying feet (the line connecting the barycenter of each foot), hip, trunk, shoulders, and head directions. This particular feature is inspired by body postures typical of classical paintings and sculptures where such angles are exploited to express postural tension.
Impulsivity	Movement which is sudden and not prepared by antagonist's muscles.
Fragility	Is a continuous re-planning of the movement, interruptions are unpredictable, and not rhythmic / non-repetitive.
Fluidity	A fluid movement ([Alborno et al., 2016]) is smooth and coordinated (e.g., a wavelike propagation of energy through body joints).

Coordination	Whether a movement is made of synchronized components (e.g., Synchronization of limbs to operate a body at the unison).
Lightness	How gravity influences a movement, e.g., based on the relationship between vertical and horizontal components of acceleration. Heavy and light movements may accentuate in different ways different parts of the performance and make a performance more expressive. This is computed from acceleration obtained either from motion capture data or from accelerometers.
Suddenness	Rapid changes of velocity in a movement. Similarly to Lightness a performance may require both quick and sudden movement strokes and continuous and sustained ones. Suddenness too may give accents to specific parts of the performance and convey expressivity. This is computed from velocity obtained either from motion capture data or from video. In case of video processing, velocity may be possibly replaced by Motion Index.
Sway	Back and forth and left and right oscillation of the body (especially of the trunk). This was already deemed important in analysis of music performance with other music instruments (e.g., see [Davidson, 2012], for a recent study). Oscillation of the head in a duo of violin players, captured by a camera hung to the ceiling, was already used for analysis of synchronization between the players [Varni et al., 2010]. Body Sway is computed from Postural Learning (both from motion capture and from video).

Table 3.3: Mid-level features, maps and shapes

Postural features from motion capture	
Name	**Description**
Openness	A measure of the extent at which body segments/joints are close or far from the barycenter. It is computed from the distance from the barycenter of each 3D tracked point. In case video input only is available, an approximation of this feature may still be computed from the bounding rectangle and the area of the player's silhouette.
Balance	A measure of the extent at which a posture is stable. This is computed by projecting the barycentre of the body to the floor and by measuring the distance of the projection from each foot.
Leaning	Leaning of the posture on the frontal (right / left) and sagittal (back /forward) plane.
Tension	A measure of the amount of torsion in a posture. This is computed from the angles between the adjacent lines identifying feet (the line connecting the barycentre of each foot), hip, trunk, shoulders, and head directions. This is inspired by classical paintings and sculptures where such angles are exploited to express postural tension.
Postural features from video (RGB-D sensor or camera	
Feature Name	**Description**
Contraction Index	A measure of the extent at which a posture, represented as a body silhouette, is closed around the barycenter or expanded, e.g., with limbs far from the body [Camurri et al., 2005].
Leaning	Right/left leaning are detected, e.g., by computing an elliptical approximation of the body silhouette and by measuring the angle between the major axis of the ellipse and the horizontal axis [Camurri et al., 2005]. In case a depth map is available back/forward leaning may still be detected by comparing the depth of the lower and of the upper body, e.g, see [Mancini et al., 2014].

Table 3.2: Postural features

Name	Description
Entrainment, Engagement, Reflective Thinking, Attention, Confidence, Uncertainty, Curiosity etc.	These high-level features represent structures and concepts related to the child's internal status and emotive component. Some are related to the individual and aim at understanding his/her emotional or cognitive state while interacting with the platform (e.g., level of Attention, Engagement or Stress). Others are inter-personal, that is, result from the interaction of two or more children (Entrainment).

Table 3.4: Concepts and structures

Part II

Case studies

Chapter 4

The weDRAW Project

4.1 General overview and objectives of the project

weDRAW – "Exploiting the best sensory modality for learning arithmetic and geometrical concepts based on multisensory interactive Information and Communication Technologies and serious games", was a two years European Project, started in January 2017. The project took its moves from the renewed understanding in neuroscience research of the role of multisensory integration in the developmental age. The core of the project was the developing of a multisensory technology on which different serious games relied to exploit the best sensory modality for mathematics learning. The concepts the project addressed where the most difficult to master for primary school children, both typically developed and sensory impaired - visually impaired and dyslexic - ones. As discussed in above sections, using performing arts, particularly music and body movements, in educational settings may be used to let the user to deeply understand the learning content in a personalized way. The association between mathematics and arts is grounded on scientific evidence; the link between drawing and geometry is suggested showing that body movement associated with spatial feedback can be used to improve spatial cognition [Finocchietti et al., 2015]. Body movement is naturally associated to space cognition. The idea to associate music and arithmetic is based on the understanding that temporal sequences at the base of rhythm and music can be naturally associated with numerosity. Besides application to typical primary school children, a major aim of the project was to apply the proposed multisensory and multimodal approach to two specific populations: visually impaired and dyslexic children. The technology mapping of the body movement in multimodal feedback – audio, visual and haptic signals, should permit to exploit the link between the proprioceptive one and feedback from other modalities.

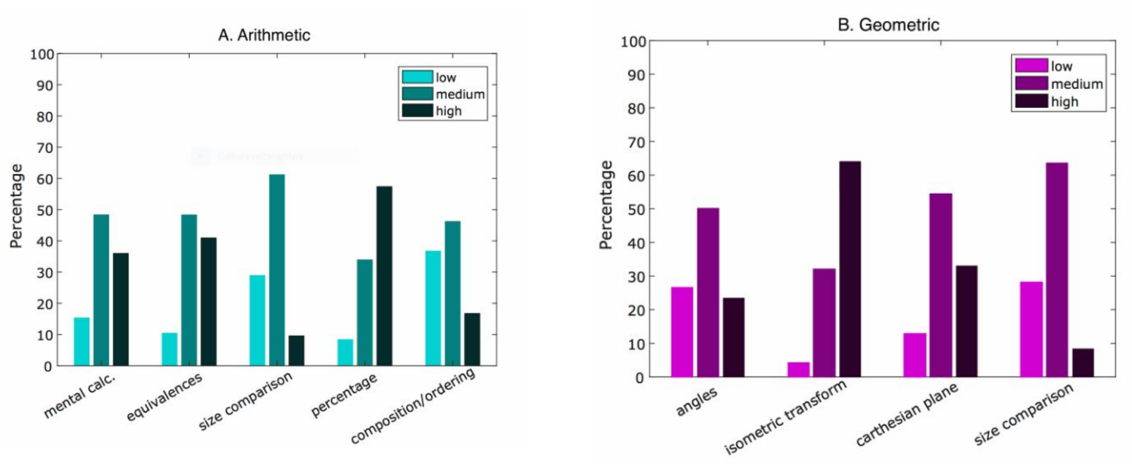


Figure 4.1: Mostly difficult arithmetical (left) and geometrical (right) concepts to learn for children aged 6-10 years old

4.2 Pedagogical Design Framework

4.2.1 Methodology

In the first year, the project has investigated the fundamental pedagogical elements for a fulfilled comprehension of the educational constraints. In order to identify the most appropriate mathematical concepts for the weDRAW project, UCL, UNIGE and IIT UVIP teams, leaded by UCL KnowledgeLab (UCLKL) researchers, collated data from teachers, both in UK and Italy, through *ad hoc* questionnaires and workshops. The concepts identified by them as most challenging for their students were then developed into design scenarios, to let teachers imagine how possible solutions might look like [Rosson and Carroll, 2009], using participatory design techniques. Teachers in UK were voluntary recruited via invitation form from the UCLKL researchers to join the project, including two mainstream primary schools (4 teachers and 2 teachers assistants) and a special school for visually impaired students with 1 teacher. Meanwhile, in Italy they were recruited, always on voluntary basis, from 4 schools collaborating within the weDRAW Project. There were: Elementary Contubernio school, International school of Genoa, Elementary school of Bolzaneto in Genova, Elementary school of Coronata. Students from the collaborating partnership schools, moreover, were recruited on a voluntary basis for psychophysics tasks and technology evaluation. A questionnaire was developed by the UCLKL team and translated into Italian by us, to be administered to the teachers, in which they were asked to quantify, according to their own experience, the difficulty children experienced understanding specific concepts, for children aged between 6 and 11 years old. Particularly, UCLKL researchers asked to:

- Select three concepts from a list, one that teachers find easy (low difficulty), one not particularly easy (medium difficulty) and one particularly difficult (high difficulty) for each age group of children to learn;
- Indicate the degree to which specific arithmetical concepts are challenging for primary school children by indicating the level of difficulty (low, medium, high) for each concept;
- Indicate the degree to which specific geometrical concepts are challenging for primary school children by indicating the level of difficulty (low, medium, high) for each concept.

The questionnaire, published for information here in the Appendix A, was administrated to 101 Italian primary school teachers (Appendix A.1). In Italy, there are 5 levels of primary education: level 1 (6-7 years old); level 2 (7-8 years old); level 3 (8-9 years old); level 4 (9-10 years old); level 5 (10-11 years old), corresponding to UK primary key stages 1 and 2. A more deep mapping of the two systems is provided in Appendix A A.2 and was given by UCLKL team. As previously mentioned in Chapter 3, the participatory design has been a central element for the methodology research followed both in this project. Participants design workshops provide critical link between technical developers and the ”on the ground” teaching practice. They formed an ongoing evaluation across the iterative development process, with each workshop building on the previous one. For example, the role of the first workshop was to provide a foundation for designs generated by teachers, trough bringing their own real-life practice challenges to the project team, and generating solution ideas, that could be developed and enhanced through collaborative brain storming with the developing teams. We managed four workshops with teachers both in Italy: 3 workshops involved teachers from the mainstream schools, and 1 workshop involved teachers and rehabilitation therapists used to work with visually impaired children. In total, 27 teachers took part, together with members of the weDRAW team, included us, IIT UVIP team and UCLKL and UCL UCLIC partners. Teachers worked in small groups for discussion and the majority of them taught children within 6-10 years of range, with one teaching children aged 12 years. The structure of workshops comprised 4 key activities;

- to introduce the weDRAW project and the team;
- to identify the key challenging mathematical concepts that might benefit most from the project;
- to identify the way these topics are currently taught in school, and the multi-sensory resources used;
- to explore some multi-sensory solution ideas in conjunction with project team to inform the underpinning of possible digital designs.

A post workshop questionnaire was also administered to enable teachers to give further information and feedback on the ideas generated, as their reflections develop within the context of their teaching.

4.2.2 Identification of the pedagogical targets

Results from this first phase of work highlighted that percentages, equivalences and mental calculation were the arithmetical concepts that teachers identified as presenting most difficulty for children to learn (see Fig. 4.1). Percentages and equivalences were more challenging for the older children, whilst mental calculation was more difficult for younger students. More in deep, UCLKL data analysis results showed that percentages only emerged as difficult at level 5 (10-11 year old) whilst equivalences emerged at levels 3-5 (8-11 years old). Mental calculation was shown as presenting the biggest challenge at level 2 (7-8 years old). It was reported that concepts of relations and representations presented a specific challenge to the youngest students, at level 1. Isometric transformations and the Cartesian plane were the geometric concepts that teachers identified as presenting the most difficulty for children to learn. Translation and rotations and Cartesian plane concepts were reported as challenging for older children, at levels 4-5 (aged 9-11) whilst orienteering was a specific challenge for the youngest students, in year 1. Spatial indicators presented a noticeable challenge (medium) for level 1 students whilst angles were difficult for students at levels 3-4 (8-10 years). Generally speaking, the results that UCLKL obtained and shared with the project consortium, highlighted that the concepts that were challenging were not consistent across the levels or ages of children. This can be explained partly because some concepts are not introduced to children until a certain age, and partly because the complexity of a concept increases as the children progress from level to level. Some concepts may be easier to understand at a basic level, but become more difficult to understand as the concept became more complex or is combined with others. Design proposals need to take into account that challenges are different within the age group being targeted. Whilst specific concepts can be identified as challenging, and perhaps suitable to teach using multisensory serious-game, they may only be relevant to a specific year within the project scope of 6-10 years of age. Counting on and counting on from non-zero starting point is problematic; if children already have a certain number, then they need to know how many more that is challenging. They typically learn to count using a song, but mapping this knowledge into objects or items is challenging: moving from the number as a symbol to what it symbolizes in terms of quantity is the core of the difficulty. Moreover, for visually impaired children moving between negative and positive numbers using a number line was also highlighted, and the idea of numbers as bigger or smaller of a certain reference quantity, the comparison between size and number, and relationship between non-corresponding sizes, e.g. children think of themselves as 6 years old ad they are bigger than their classmates who are 7 years old; and the difference between number and quantity, and the association between quantity and number are some of the troubles they found. For what is concern **fractions**, they were identified as challenging by all groups of teachers, with specific issues being raised:

- Understanding the concept of 'a whole' or the 'unit' was raised by most groups, including the visually impaired. This suggests children have difficulty with 'framing' a fraction - understanding the fraction in relation to the implied whole (if 2/3 = 20 then the whole must be 30).

- Equivalencies (a half is the same as two quarters is the same as 50%). Children remain anchored to the number quantity and are not able to evaluate it in a relative manner within the fraction. It is difficult to make the symbolic passage linking the number to the entire part of the fraction. For example 5/10 same as 2/4 same as 9/18, and the related simplification of fractions e.g. 6/9 divide by 3 to get 2/3;
- Direct and complement fractions (if you subtract 2/3 then you must have 1/3 left). Understanding that a larger number as denominator does not mean a bigger fraction.
- Fractions and portioning: for example, "if I have 20 chocolates and want to divide into 5 equal portions, how does that happen?". Teachers reported that this is generally conceptually understood, and easy to apply to regular shapes, e.g. circles (pizza, clock quadrants), squares, but hard to apply to irregular shapes.
- Division and percentages, making the link between fractions, percentages and decimals (e.g. 0.2 = 1/5 =20%). The national curriculum specifies that pupils should be taught throughout that percentages, decimals and fractions are different ways of expressing proportions [for Education, 2013]. This also extends to geometric concepts, for example linking percentages to 360° in order to calculate the angles required for pie charts. In particular for visually impaired children, a lack of practical experience of what percentages are and where they are used, leading to a challenge transferring from class to real life.

For what it concerns geometry, measurement, such as perimeters, areas and volumes in general were identified, but specific challenges were:

- **Isometric transformations**: reflection, rotation, translation, including distinguishing between terminology around transformation, translation, rotation, reflection. The terminology leads children to blend the concepts into one idea. Understanding that a figure rotated is always the same figure, is challenging. Understanding angles and their rotation, was particularly noted, specifically retention of the different angles. Children use strings, body movement, fans etc. for exploring angles. They identify them in class but cannot recognise an obtuse compared to a right angle when they come back: once lesson over the ideas seem hard for them to retain (8-11 years).
- **Symmetry**: visualising lines of symmetry, and understanding the properties of irregular shapes. Symmetry is considered easier for simple shapes like circles and triangles but children struggle once the shape becomes more complex; and when the point of rotation is remote from the shape it is hard for students to visualise it.
- **3D shapes**: understanding the third dimension of a cube or other 3D shapes
- **Cartesian plane**: understanding the difference between 3D objects that they experience in the real world and their 2D representation. Being able to understand how coordinates can be used to represent a shape like a rectangle.

For the visually impaired, the following geometrical challenges were also noted:

- **Idea of shape**: Understanding the relationship between what this is the shape and what is not the shape. The children need to understand the beginning and end of the shape; they need reference points on a spatial and temporal continuum

Looking for initial solutions to inform the technological design, we engaged teachers in group brainstorming to generate the seeds for design ideas, during a series of workshops done in Casa Paganini InfoMus-Lab. To ensure the goodness of the results, it was critical to built on current teaching practice, focusing on the particular conceptual challenges identified, fostering real world experience wherever possible, considering also cross-concept aspects. Some concepts, such as constructing physical shapes using paper to demonstrate nets or folding paper to explore symmetry, were thought to lend themselves more naturally to multimodal approaches than others, such as line number. Colour was recognized as a visual resource, but audio or tactile resources were perceived as less commonly used in classroom. Some activities described by the teachers, such as manipulating paper into 3D shapes, folding to find lines of symmetry or using a trundle wheel suggest that there may be an unrecognised tactile or audio aspect. Three general factors emerged from the workshops: (i) anything that will give learners a different experience, whether it is to help them remember or to engage, could help; (ii) that we make use of prior knowledge and link concepts to help build on week by week teaching. Day-to-day physical real world experience is important and not every child has the same access to this, for example, children whose parents regularly involve them in activities like cooking or shopping have a better understanding of measurement; and (iii) we need to recognize that classrooms are crowded with different resources, so what we develop must be useful, and where possible valuable for more than one task or concept.

4.2.3 Initial Design Ideas: Arithmetic

Five ideas relating to the ‘number line’, number and quantity, and ‘counting on’ were put forward:

- Using music or a specific sound for each number, thus children know that “each number has a sound, 7 always sound like this”. Thus, ‘counting’ sounds different depending on where you start along the number line, supporting the notion of ‘counting on’.
- Have a pathway of children standing in a row, each of them representing a number from one upward. One child moves along pathway, made up of children who represent 1, 2, 3, and collect a child as they go, such that two children move together to the third child. In this way, they can ‘see’ that the number relates to a quantity, not just a word. The increasing quantity up the line could also be associated with audio, e.g. volume.

- Number line that looks like a staircase, so giving the ideas that quantity is like something that is growing: walking up maps to increasing quantity. There could be a sound for going up and one for going down to direct the children, and that could map to addition and subtraction. The number could also be mapped to a number of beeps children hear on each step; and/or to tactile feedback either on the 'stairs' or through an object held by the child;
- A roll out carpet with number segments, which elicits audio as the child walks along the carpet cells (e.g. scales). There could be a 'pocket' on each cell with tactile objects inside each pocket, so that the quantity of objects maps to the numbers written on the number line. Teachers also suggested a voice synthesizer that 'speaks' the number as you stand on it; or a nursery rhyme/ relevant song for each number that is associated with the number line e.g. like "12 days of Xmas" (maps to ideas in number 1);
- Series of speakers, so as you move the past each speaker then could get audio feedback. This could be increase in amount – cumulative or individual representation of quantity.

For what it concerns framing fractions or Cartesian plane, where students need to understand related concepts, such as "unit" and "whole", one suggestion was to get children to jump, where the 'jump' creates the unit. This needs to then be divided into smaller units to represent the fraction. One other was to fill the space between jumps with music/audio that, for example, represents thirds using a rhythm with 3 beats. A similar augmentation could be implemented using haptic/tactile feedback. One group also suggested using spatialized sounds and lights to understand fractions. A key aspect of using music or haptic rhythm in this way is to meaningfully frame the whole beat in order to define the fractions within that frame. One option might be to have a group of children representing a 'whole', with each child representing a part of that whole. Children could jump according to which fraction of the whole they are asked to demonstrate. Some key questions are: how do we define arithmetic using music e.g. pitch or volume, can children distinguish between different sounds, especially for numbers that are quite close? Or how do we represent very big numbers?

4.2.4 Initial Design Ideas: Geometry

Moving forward to geometry concepts, some ideas involved:

- Using the virtual space or (tangibles) digitally augmented objects to explore shape and transformation. For example, instead of a square based pyramid children build a real Egyptian pyramid, using 'tangible' shapes to build a city.
- Issue with cube and understanding the different faces, different angles and inside corner/ 'blind' corner, using the haptic platform for the 3D shape understanding. For example, there could be a transparent cube with a different colour for each face, so can see where

you are in the cube; cube can be opened flat and then ‘folded’ back up again. This was thought to help by being able to open the cube and see inside it and see the blind angle; having different colours, and the transparency will help to see what is in the back (the hidden corner). There could also be different shapes that have the same volume; children can open the surfaces of the shapes and see the different areas; and move from the 3D to the 2D representations of the shapes.

- Using the body and its symmetry for teaching angles and shapes: for example, four children can compose a cube with their body, by opening the arms to create an angle and rotate or using elastic strips to connect different body parts (e.g. by connecting a leg with an arm) to simulate a shape. For triangles children can use their body and projecting arms to the floor in a triangle, or using their feet. These bodily movements could be linked to sounds that relate to the type of triangle, or the different sides of the object, or the different angles. This activity could be done with more than one child together.

4.3 Psychophysics evaluation of sensory modalities

The weDRAW project proposed an embodied and enactive approach for learning. Enactive knowledge is not simply multisensory mediated knowledge, but is knowledge stored in the form of motor responses acquired by the act of doing. For this reason, during the first year of project a great effort was devoted to study the motor and sensory modalities both of typically developed and sensory impaired children between 6 and 10 years old. These tasks were mainly developed and carried on by the weDRAW coordinator IIT UVIP group, supported by our research team, together with the psychology partner from Trinity College of Dublin. The tests The experiments presented in this section, therefore, have not been a direct responsibility of the author, but a fundamental step towards the development of the technologies that will be presented in the next chapter 5. We have chosen to include the description of some psychophysical experiments, in which our contribution was limited to a participation in their design and data collection, because they are the fundamental element on which the choice of how to use movement and other sensory channels for learning specific concepts rested.

The objectives of the test definition and evaluation of mobility skills in children were:

- To develop a battery of tests that assess motor abilities and multisensory integration in children aged between 6 and 10 years old.
- To administer these tests in sighted children from 6-10 years of age to gain important insights for the design of learning technologies for primary school children.

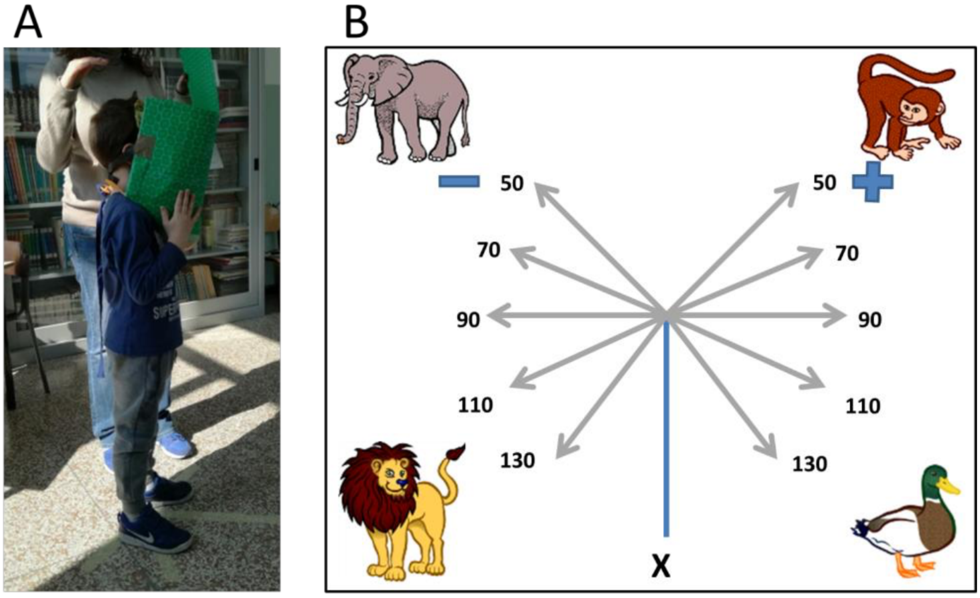
The main goal of the evaluation of mobility skills in children was to provide inputs to develop suitable technologies that can be used by children at different ages. Its this central role that convinced the author to present some of these tests in this section, to highlight, one more time, the strength relation to technology development. We then actively contributed firstly reviewing the technologies currently available and secondly contributing in designing a set of tests to evaluate mobility skills in children between 6 and 10 years of age. For example, weDRAW IIT UVIP coordinator identified the level of motor fluency and shape recognition capability of children at different ages in order to provide inputs about which kind of behaviour and feedback the technology could exploit. The mobility skills tested were wide-ranging, including locomotion abilities as well as dexterity using movement of the arm, hands and fingers (haptic perception). This versatility allowed us to test the more appropriate embodiment of mathematical concepts across different age groups, that can be implemented in technology. Understanding the developmental rate of motor skills was indeed crucial for the design of child-appropriate technological solutions which will consider full-body movement sensors.

4.3.1 Development of angle discrimination using whole-body movements

Capacity of discriminating angles of different degrees is acquired in primary school and prepares for learning geometrical concepts of increasing complexity, e.g. trigonometry. One of the first steps is learning how to differentiate between acute and obtuse angles. In order to accomplish this task it is important to understand the concept of right angle (i.e. 90°), as this represents the differentiating threshold between acute and obtuse angles - that is, in psychophysical terms, the point of subjective equality (PSE) on the acute-obtuse scale. Mostly, research investigating this capability focused on visual and haptic cues to perceive angles but not many employed active locomotion as means to perceive angles of different aperture. A recent article by Jetzschke and colleagues [Jetzschke et al., 2016] has shown the presence of idiosyncratic biases in adults when requested to report whether they perceive to have turned more or less than 90° in a walking task. In this experiment IIT UVIP team used a similar protocol in order to investigate the presence of inherent biases in perceiving the turning angle of 90° in children. Forty-two children (age range: 6-11 y.o.) were recruited.

Participants were blindfolded and additionally provided with a customized silicon bib to make sure they could not see their steps and the floor (see Fig. 4.2 A). The experimenter guided them along two path segments (see Fig.4.2 B) - the first one 225 cm long and the second one 120 cm long. The angle between the segments varied on a trial-by-trial basis in the range between $\pm 70°$ and $\pm 130°$ with a step size of 20°. Each stimulus level was repeated 5 times, leading to 25 trials per block of trials; each block lasted $\sim$30 minutes. Each subject was presented with the same order of trials randomized once across trials. Movements to the right (positive sign) and to the left (negative signs) were performed in separate blocks of trials and in separate days in order to assure participants' completion of the task. Subjects were guided by the experimenter along the two segments and at the end of the second segment they were asked to indicate towards which direction they were conducted. Since children might have difficulties in understating the concept

Figure 4.2: Example subject (A) and illustration of the experimental procedure (B)



of angles and especially the right angle, we employed visual landmarks positioned on the floor to be used as a reference for their response (see Fig.4.2 B). Before each trial, the subject was asked to remove the blindfold and look at the 4 visual landmarks. After the guided movement, subjects were asked to report whether they arrived closer to one landmark or the other. For each block of trials subjects were forced to use the landmarks of the target side, either to the right or to the left. The landmarks were drawings of animals which were part of a story told to the subject in order to make the experiment enjoyable: each participant was told that they were going to explore a forest by being hidden inside a bush (represented by the customized silicon bib); their task was to understand whether they went closer to an animal or the other and report the response to the experimenter. After each trial, the experimenter accompanied the subject back to the start position by making a random path in order to avoid potential feedbacks about performance in the task. Fig.4.2 showed averaged biases and standard deviation (SD) by grouping children according to three age groups: 6-7, 8-9 and 10-11 y.o. Biases were mostly negative both for left and right sides and for all ages. The oldest group of participants (10-11 y.o.) showed a tendency of bias direction to the opposite direction with less pronounced biases for the left compared to the right angles. Regarding precision, it can be seen that the younger the children group, the higher was the SD, thus indicating lower precision at this developmental stage as previously observed in other studies investigating multisensory perception across development [De Vrijer et al., 2008]. As in the work of Jetzschke and colleagues [Jetzschke et al., 2016], IIT UVIP team tested whether the perceived angle of 90° to the right and to the left are combined in a bias for straight ahead. They did not observe any significant correlation for all age groups

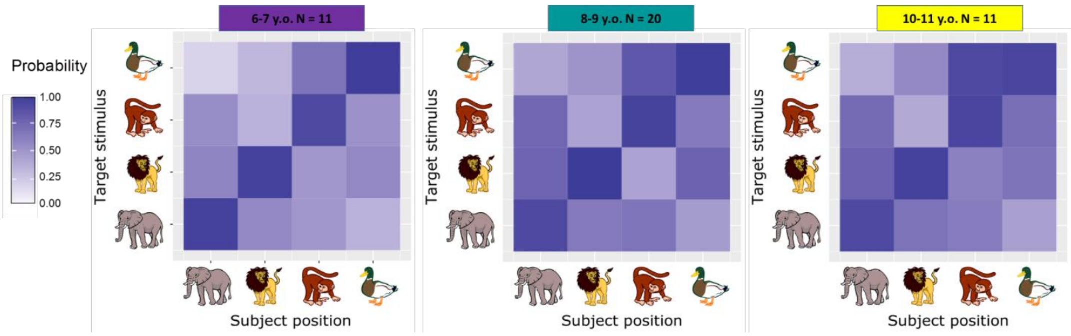
(Pearson correlation: 6-7 y.o., R = 0.42, p = 0.19; 8-9 y.o., R = -0.29, p = 0.2; 10-11 y.o., R = -0.21, p = 0.52.), thus indicating that children do not integrate biases for the left and right side in a bias for straight ahead.

This experiment has a profound impact on the ongoing weDRAW project as well as on my research, since it provided the scientific basis for the development of one of the weDRAW activity of which the author was in charge, *the AngleShape Game*, that will be presented in next paragraphs.

4.3.2 Self-motion and visual landmarks

Self-motion helps in defining visual landmark positions. When moving in space, it is fundamental to maintain a reliable spatial representation of the surroundings in order to successfully orient in space. Such a property of spatial navigation is usually guaranteed by vision as it provides constant monitoring of the environment [Nardini et al., 2008]. Nevertheless, understanding how well locomotion in absence of vision can guarantee functional perception of the environment is useful to understand how motor and sensorimotor processes can be used to learn coordinates and object position relative to self-motion. This capacity is supported by spatially updating our position in space with regard to external landmarks. Several studies investigated these properties of spatial representation and navigation by focusing on the difference between ego- and allocentric reference frames [Loomis, 2014]. In this experiment, the aim was to disclose such properties of perception by focusing on children and movements along paths that describe angles. IIT UVIP team and us tough that this aspect could help understanding whether children can retain spatial information about the position of other objects in the environment after moving through space. This experiment was conducted as a second task to be performed in 2 stimulus repetitions during the experiment described in the previous paragraph. In 14 pre-selected trials per block, subjects were requested to indicate with their index finger the position of 4 landmarks signaling the 4 quadrants describing the space where the subject moved (see Fig. 4.2 B). As previously described, subjects saw the landmarks before performing each movement, then they were asked to point at each landmark following a pre-fixed order randomized across trials at the end of each movement. The order was the same for each subject. This means that after having provided the answer relative to their movement in space (i.e. closer to which landmark they arrived), the experimenter asked them to point at the landmarks one by one and recorded the response by writing toward which quadrant subject finger was pointing. The results, obtained by IIT UVIP unit, are shown in Fig.4.3. At all ages children are capable of successfully indicating the location where they arrived with a probability of giving the correct answer above 85%. When children instead are located in positions that do not correspond to the target landmark (e.g. they are located close to the "elephant" and are asked to indicate the "duck" landmark), their accuracy decreases. Interestingly, it can be noted that accuracy decreases even more for the youngest group of children (i.e. 6-7 y.o.) compared to the older participants (i.e. 8-9 and 10-11 y.o.). This result indicated

Figure 4.3: Matrices of confusion: association between subject position and responses. Each matrix represents the results for each children group defined by age. The level of probability is associated with the coloured scale on the left: darker colour means high probability and lighter colour low probability of indicating the specific target landmark.

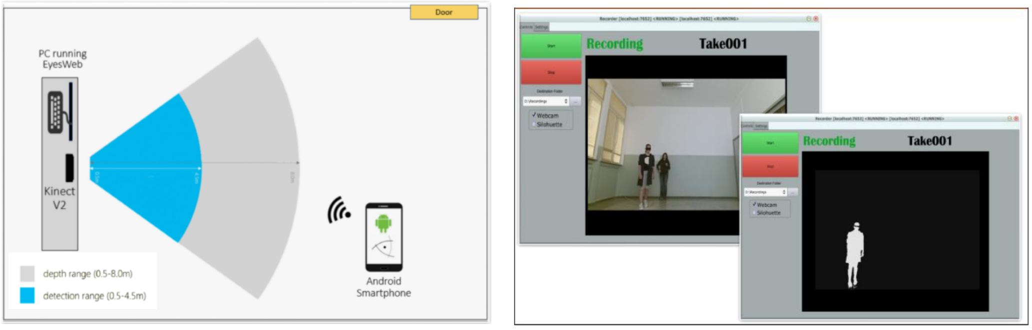


that spatial representation and especially coordinates can be distorted after locomotion without vision. This methodology allowed us to understand whether spatial updating was successfully achieved by children or not. Such capability seemed to differ depending on the age group but other factors as school level and individual differences may have a role in distortion in the spatial representation of the surroundings. This experiment aimed to disclose the influence of these factors thus providing the scientific base for the technological platform we developed, in which we used locomotion to teach coordinates in space (e.g. the Cartesian Garden, developed by the partner of Ignition Factory and for which we designed the sonification strategies suitable both for typically developed and visually impaired children).

4.3.3 Triangle completion experiment using whole-body movements

Much research has employed navigational and path integration tasks in order to study motor skills [Loomis et al., 2001, Philbeck et al., 2001]. The triangle completion task is one of the most common and examines how well individuals can update their position in space by actively moving without visual feedback. In the current experiment, we use the triangle completion task to examine spatial updating in children following turns of different degrees. Forty children (age range: 6-11 y.o.) were recruited. The trajectory were recorded by means of a Kinect sensor, that is a motion sensing input device by Microsoft, and the EyesWeb XMI platform [Volpe et al., 2016], presented in Chapter 3. The system developed to support phsycophysics tests during the project was a portable body movement capture setup, suitable to be installed in different classrooms. The setup consisted of a computer running the EyesWeb XMI software and a Microsoft Kinect v2 depth sensor connected to the computer through a router. This equipment was placed in a corner of the experiment room and a smartphone based remote controller was provided to help the operator to control the recording of the child's movements without interfering with the exper-

Figure 4.4: The user interface of the EyesWeb recording application



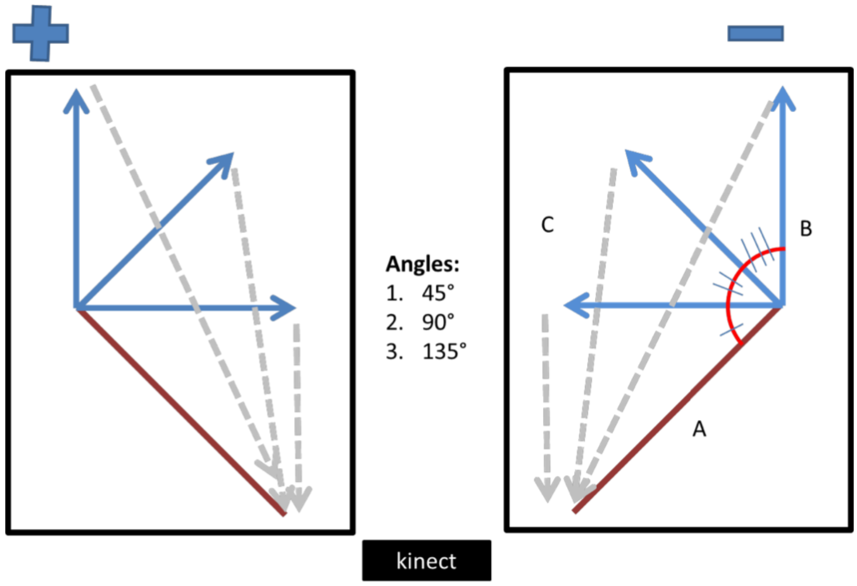
iment. The remote control was implemented as an Android application which provided access to the main functions (e.g, start or stop recordings) of the EyesWeb XMI software and allows the tuning of several parameters (e.g, subject name, experiment type, etc.). The Microsoft's Kinect v2, interfaced with the main recording module in the EyesWeb XMI software for this particular purpose, provides access to the following data streams, at 30 frames per second:

- Video (1920x1080 resolution)
- Infrared (512x424 resolution)
- Depth images (512x424 resolution)
- Skeleton information (25 full body joints per tracked user)

The available tracking volume (range of depth) was between 0.5 to 4.5 meters. Fig. 4.4 (left) showed a schematic representation of the setup, whereas on right, it presented the user interface of the main EyesWeb recording application.

Participants were asked to perform a path integration task: after being guided along two legs of a triangle (A and B, see Fig.4.5), they were asked to go back to the start position by completing the triangle, that is along the third leg (C). We manipulated the angle between A and B triangle legs to be either 45°, 90° or 135° to the right (positive sign) or to the left (negative sign). Each stimulus level was repeated 4 times, leading to 24 trials per block. The experimenter guided participants in walking along the two legs of the triangle: turns were made by telling subjects to turn and gently pushing them towards the target direction with experimenter's hands on subjects shoulders. When the experimenter and participant reached the end point of the second triangle leg, the experimenter moved away from the participant and asked them to go back to the start position with no help from the experimenter. In order to make the task more enjoyable, children were told that they were walking through a wild forest and they had to go back to the start position (indicated as "home") by doing the shortest path and maintaining constant velocity.

Figure 4.5: **Illustration of paths performed during the experiment.** Positive and negative signs refer to turns to the right (positive sign) and to the left (negative sign). The red lines represent the first segment (200 cm) walked by subject and experimenter. The blue arrows represent the range of possible segments (150 cm) walked by subject and experimenter. Gray arrows represent the ideal path walked by the subject without the experimenter. Behind the start position there is the Kinect system recording subject performance.



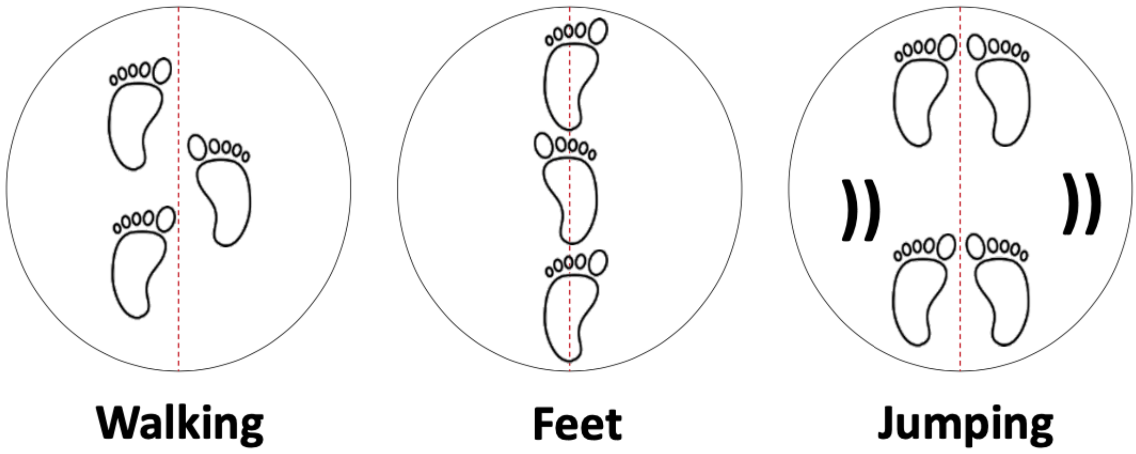
Before starting the experiment, participants were informed that they had to complete the triangle by walking along the third leg of the triangle; in order to make this point clear, the experimenter showed them a picture of a triangle, the path they were going to do with and the one they will do without experimenter guidance. Participants were blindfolded at the beginning of each block of trials by means of swimming goggles occluded with black tape and positioned over a blindfold that covered subjects eyes.

A more comprehensive analysis of this data and the results obtained are presented in an article, write with the colleagues of IIT UVIP team, and actually under revision.

4.3.4 Space Segmentation

The aim of this experiment was to assess the ability to segment a distance in space by using the body. Specifically, it was tested if children correctly perceive the length of a path when asked to employ different strategies to segment the path. IIT UVIP team, in collaboration with us, developed a space segmentation task assessing the ability of children to perceive and segment distance in space (Fig. 4.6). During the demonstration phase, children were guided by the

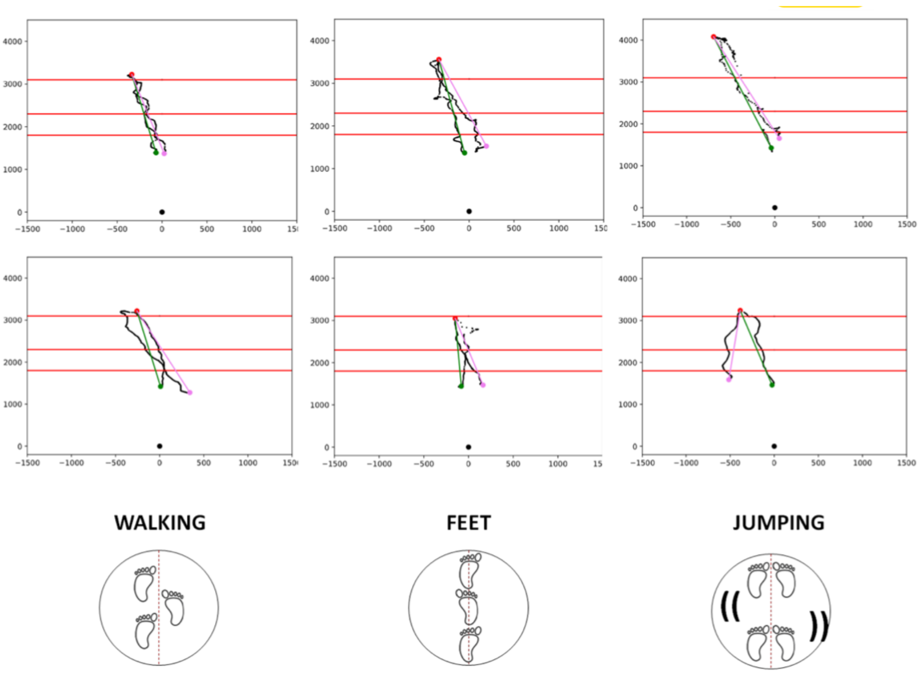
Figure 4.6: **Experimental protocol of the space segmentation task.** Children were asked to reproduce a distance in space by walking, putting feet one after the other and jumping.



experimenter along a path of a specific length: the path could be short (70 cm), medium (120 cm) or long (200 cm). Immediately after during the reproduction phase, children were asked to repeat the same path under three conditions: by walking at their own pace (walking condition, Fig. 4.6), by walking putting one foot in front of the other (feet condition, Fig. 4.6) and by jumping (jumping condition, Fig. 4.6).

The length of the path was indicated by coloured tape on the floor. Each reproduction condition was intended to segment the space included in the path previously experienced by the child: contrarily to the walking condition in which children are not expected to use different strategy to understand the length of the path, in the second and third condition children are supposed to mentally rescale the distance previously perceived according to the reproduction condition requested. We assessed the performance of a sighted group of thirty-five children (age range: 6-11 y.o.). In total each child performed 27 trials: each distance (short, medium, long) was repeated three times for each condition (walking, crossing feet, jumping). Participants were blindfolded at the beginning of each block of trials by means of swimming goggles occluded with black tape and positioned over a blindfold that covered subjects' eyes. To measure spatial accuracy, the trajectories performed by children were recorded by means of the Kinect with the EyesWeb recording App mentioned above. As an initial measure of spatial accuracy, we computed the localization error measured as the distance in cm between the correct final position of the path and the perceived final position of the path indicated by the child.

Figure 4.7: **Representation of individual subjects' performance in the space segmentation task.** The graphs represent the individual performance of a 6-years old child (upper panel) and a 10-years-old child (bottom panel) in the reproduction of a long path (200 cm) corresponding to the first red line from the top. The starting position of the child was set at 110 cm from the spatial position in which the Kinect was positioned (black dot) therefore the long path measured 3100 mm. Black lines represent the trajectory performed by the child, the pink and green lines represent the calculated regression of the whole path, the red dot represent the final position of the path performed by the child. The distance between the red dot and the first red line from the top has been considered as mean localization error (in mm). As can be seen, while the performance of the 6-y.o. child slightly deteriorates in the feet and jumping conditions, the performance of the 10-y.o. child is not affected by the reproduction condition.



4.4 Discovering the best sensory modalities to teach mathematical concepts

Previous research studies showed that crucial arithmetic concepts (e.g., counting, measuring distances, and so on) can be heavily grounded in other modalities than visual, such as auditory and tactile modalities. For example, the relationship between critical components of music (pitch, melody, tempo, rhythm, and harmony) and numbers and fractions has been recognised since ancient times [Mall et al., 2016]. In recognition of these associations, recent studies suggest a growing interest in an integrated approach to the teaching and learning of music and mathematics [Viladot et al., 2018]. The paradigms used for testing were grounded on a body of scientific evidence that enabled us to identify the most optimal sensory modality through which to teach specific concepts. This allowed the child to acquire simple associations with arithmetic concepts (e.g., few vs. many, counting, sum, subtraction, and so on) that are easy to understand. Based on the results of the tests reported here, we developed a personalized approach that used the preferred learning modality at the individual age groups (e.g., for learning shape at one age children could prefer to use the visual signal while at another age might rely on tactile signals). These preferential signals for each age group then allowed us to exploit different sensory modules for different children. Since the teacher mediated, through the entire project, this process, we used questionnaires to collect teachers' feedback and opinion regarding the best learning modality. The teachers' opinion on the best modality, based on their classroom experiences, acted as a starting point for the validation process based on the psychophysical tests. Besides arithmetic, also geometry is at the foundation of Science, Technology, Engineering, and Mathematics (STEM). Application of geometry in drawing and painting is widespread in all ages (since the Renaissance studies on perspective to contemporary art, e.g., cubism). Although both the concepts has always being mostly based on the visual sensory channel, both arithmetic and geometry are largely taught from symbolic or iconic approaches. In symbolic learning, knowledge is stored as words, mathematical symbols, or other symbol systems; whereas in the iconic stage, knowledge is stored in the form of visual images, such as diagrams and illustrations that may also accompany verbal information. Changing the perceptual channel from the visual one to the most suited one, or adding further perceptual channels, does not necessarily change the concept that is being learned, although it will change the way in which it is learned. The scope of the definition of the best modality for learning arithmetical concepts in children between 6 and 10 years of age was to develop new technological solutions within the weDRAW project. In previous studies it has been observed that children start to integrate multisensory information (such as vision and touch) only after 8-10 years of age [Gori et al., 2008], so enduring the last years of the primary school. Before this age, performance is not optimal and the most robust sensory modality teaches (calibrates) the others. For example, it was observed that children use the tactile modality to perceive the size of objects whereas the visual signal is used to perceive their orientation. This suggests that specific sensory modalities can be more suitable than others to convey specific information and hence to teach specific concepts.

4.4.1 Evaluation of the best modality to teach arithmetic

4.4.1.1 Number line: the influence of sound pitch on length perception

The goal of this work was to investigate the association between pitch and length perception in order to gain insights about the role of sounds and auditory perception in mathematical learning. Specifically, the hypothesis it was investigated was if children are influenced by the pitch of a sound when asked to judge the length of a path, given that pitch is known to generally influence size perception [**?**]. The task required children to perceive and segment a distance in space by walking and by jumping (condition two and three of the experimental protocol). During the demonstration phase, children were guided by the experimenter along a path that could be short (70 cm) or long (200 cm). During the reproduction phase, children were asked to repeat the path by segmenting space. Before or during the demonstration phase a second experimenter switched a sound on: the sound could be a low pitch tone (400 Hz) or a high pitch tone (1000 Hz). We used two modes of sound presentation. In the first case (mode one), the sound was turned on for three seconds before the beginning of the demonstration phase. In the second case (mode two), the sound was turned on for the entire duration of the path. The aim of this experiment was to assess if pitch can also influence the ability to perceive and segment the length of a path. Moreover, in mode one, the duration of the sound can provide cues about the length of the path in advance. Thus, by comparing the two scenarios for sound presentation, it is possible to gain insight into the role played by pitch and duration of sound on the perception of length. We assessed the performance of a sighted group of sixteen children (age range: 6-11 y.o.). In total each child performed 16 trials: each distance (short, long) was repeated 4 times for each condition (crossing feet, jumping) and 2 times for each sound mode (before or during the demonstration phase). Participants were blindfolded at the beginning of each block of trials by means of swimming goggles occluded with black tape and positioned over a blindfold that covered the subject's eyes. To measure spatial accuracy, the trajectories performed by children were recorded using the Kinect sensor with the EyesWeb recording App presented above.

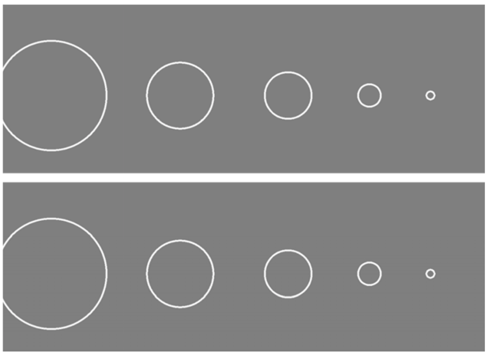
4.4.1.2 Fractions with the body: upper-lower part

In this experiment we aimed to investigate how well upper and lower body parts are perceived by children. Body representation can change depending on the developmental stage. In this context, it is of interest to quantify the biases, if any, in perceiving arm aperture (intended as the distance between the hands on a horizontal plane) relative to legs aperture (intended as the distance between the feet on the ground). This aspect was then be employed to understand how arm and leg apertures can be used to convey information about size or more general quantities. We collected preliminary data on 4 children (age: 7-10 y.o.) who performed the task. We asked participants to perform a 2AFC (two-alternative forced choice) task where they had to discriminate between hands and feet distance. At the beginning of the experiment, participants were blindfolded by means of swimming goggles occluded with black tape and positioned over a blindfold that covered subjects eyes. In order to track hand and feet distance we used the Kinect and EyesWeb recording App. In the calibration phase, participants were asked to extend their arms and their legs to reach the maximum distance between hands and feet, respectively, and these lengths were recorded as the maximum reachable by the subjects. Regarding the maximum distance between the feet subjects were asked to reach a distance that could still be comfortable with no risk of losing their balance. After the calibration phase, subjects started the experimental block of trials: on each trial, they were asked to extend their arms and their legs until they reached the target extension signalled by a sound from the computer speaker. Firstly, the sound feedback told subjects when to stop extending their arm extension as they gradually stretched them out, then the same procedure was done for the legs by using the same sound. In order to assure that the procedure was correctly performed, the experimenter was constantly providing feedback to the subject in addition to the sound feedback. At this point, the experimenter asked them to indicate the body part separated by the greater distance – the arms or the legs. Target distances were in the range between 0.1 and 0.8 arms and legs aperture (where 1 means the maximum aperture) and they were compared with an aperture of 0.5. Each aperture comparison was repeated 3 times leading to 48 trials in total. Preliminary results indicate a mismatch between hands and feet distance as in several cases children reported either hands or feet distance to be larger than the other also in those trials where the actual discrepancy was in the opposite direction. Although these results are still preliminary, recording of such a behavioral tendency hints to possible intrinsic biases in discriminating the upper and lower part of the body. Future extension of these results will bring information about the sensorimotor processes behind upper and lower body parts perception. This information was fundamental for developing a full-body serious-game, that will be presented in next sections, that took advantage of the discrepancy between upper and lower part of the body to teach fractions.

4.4.2 Evaluation of the best modality to teach geometry

The scope of the definition of the best modality for learning geometric concepts in children between 6 and 10 years of age was to improve the development of technological solutions within the weDRAW project. The definition of the best sensory modality that could be applied to the acquisition of each specific geometrical concept was crucial to provide inputs for the development of technology targeting multiple sensory modalities such as vision, audition and tactile/haptic signals for learning specific concepts (e.g., touch/haptics for geometry). The weDRAW consortium (led by UCLKL) identified a number of basic geometric concepts from primary school curricula that are based on bodily forms of interaction and map well to the aims of the project. Below is the list of particular geometric concepts identified by teachers and specific challenges. Of particular note was challenges relating to 'measurement' (i.e. understanding perimeters, areas and volumes). **Trasformations: -Isometric transformations:** reflection, rotation, translation, including distinguishing between terminology around transformation, translation, rotation,

Figure 4.8: Stimuli used in the identification task.

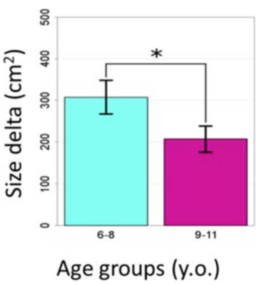


reflection. -**Symmetry:** visualising lines of symmetry, and understanding the properties of irregular shapes. Shapes: -**Cartesian plane:** understanding the difference between 3D objects in the real world and their 2D representations. -**3D shapes:** understanding the third dimension of a cube or other 3D shapes. The investigation between sensory modalities and geometry concepts was quite entirely conducted by IIT UVIP and Trinity College teams. Despite that, here some studies are anyway presented, since they represent the psychophysical basis of the technological solutions we designed and present in this Chapter.

4.4.2.1 Shapes: pitch and size, audio-visual cross-modal interaction

Multisensory association between sensory features of different nature can be observed in phenomena known as cross-modal correspondences [Parise, 2016, Spence, 2011].

Figure 4.9: Size delta is the difference between response to 250 Hz and response to 5000 Hz.



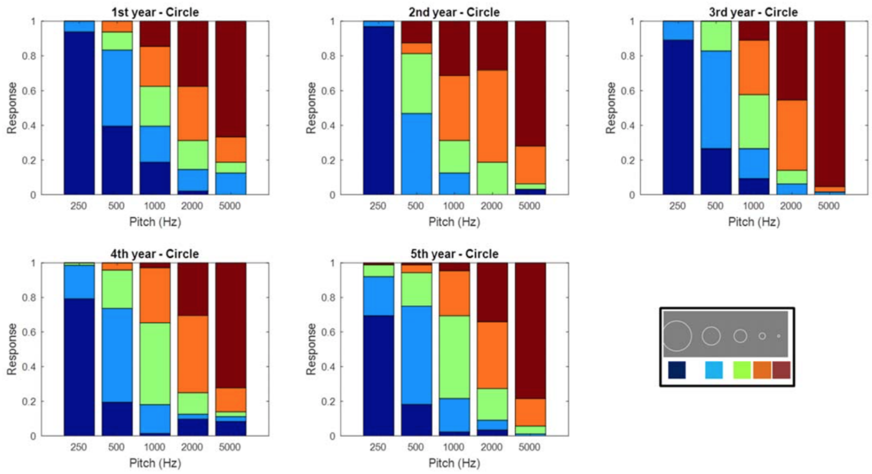
Correspondences of this kind are present in everyday life: one example is the association between auditory pitch and object's size that can be defined via visual or haptic perception. In the literature, there are several evidences of correspondences of this kind [Parise, 2016], [Spence, 2011]. Overall, relative associations between features belonging to different sensory modalities can be observed: a particular feature in a sensory domain (e.g. visual size) and another feature in another sensory modality (e.g. auditory pitch) can show a congruent (small objects and high pitch) or incongruent (small objects and low pitch) relationship. In particular, it has been shown that perceived size of visual objects can be modulated by additional sensory stimulation transmitted through other sensory modalities (e.g. auditory) [Gallace and Spence, 2006, Parise and Spence, 2009]. For example, when visual objects are presented together with low or high pitch sounds their perceived size can be biased: low pitch sounds can increase perceived object's size, while high pitch sounds can decrease perceived size [Jaekl et al., 2012, Pooresmaeili et al., 2013].

Previous studies investigated this relationship in children but findings so far are not conclusive, some show that children do not gain this association until 11 years of age [Marks et al., 1987], others show audio-visual and tactile-visual cross-modal correspondences in preschoolers when judging motion stimuli [Nava et al., 2016]. In young children, cross-modal correspondences relative to size have been shown when loudness is used as an auditory feature [Smith and Sera, 1992], but less is known about the effect of pitch. In order to test whether the association between sound pitch and size is present in scholar children and if it changes depending on age, school level and individual differences, IIT UVIP team, supported by us, ran two experiments focusing on this visual-auditory interaction relative to size. 71 children, ranged in age from 6 to 11 years, were tested in two different tasks: reproduction and identification. In both tasks, auditory stimuli were pure tones of different pitch levels (250, 500, 1000, 2000, 5000 Hz) and were presented by one loudspeaker positioned in front of the subject. The presentation order was randomized once and this fixed order was used for all subjects. Responses were taken with a touch screen monitor positioned in front of the subject. In the reproduction task, after listening to the sound for 2 seconds, subjects were asked to draw a circle "as big as the sound was". Additionally, it was asked them to rate from 1 to 10 the size of drawn shape (from very small to very large) after each drawing. In the identification task, participants were presented with the sound stimulus and were asked to indicate the shape that better represents the listened sound. Subjects were presented with two identical arrays of 5 visual objects one above the other (see Fig. 4.8). The objects composing the arrays were circles of different sizes.

Results for the reproduction task showed that children of every age associate large drawings with low pitch and small drawings with high pitch sounds thus indicating that such association is present already at 6 years. Moreover, the reported numerical rate correlates with the size of the drawn shape and confirms that children understand drawings of different sizes. Interestingly, the younger the child, the greater was the discrepancy between responses for low and high pitch stimuli.

The results showed that discrepancy between reproduction responses for low VS high pitch sounds depended on the developmental stage, as it is greater in the youngest (6-8 y.o) compared to the oldest children (9-11 y.o.). Results from the identification task confirmed that children of every school level showed such cross-modal correspondences between auditory pitch and visually defined size. The analysis of variance (ANOVA) showed that there was an effect of the sound pitch level on choosing the circle that corresponds to the sound pitch whereas there was no influence of the school level ($F(4,1675) = 0$, $p = 1$) nor of the interaction of the two factors These results showed that this test could be helpful in understanding and quantifying multisensory processes subtending size perception for every age of interest. They also suggest that sensory

Figure 4.10: **Results identification task.** Each plot indicates the school level. Each colored bar represents the proportion of responses for each auditory stimulus; the five different colors refer to the 5 circle options.



modalities other than vision could be used in learning size comparison across development.

4.4.2.2 Shapes: Angles pitch and aperture, audio-visual cross-modal interaction

In this study, it was investigated if sound frequency can be used to understand angle aperture.

Figure 4.11: Stimuli used in the identification task. From the left, the angles were: 100°, 60°, 40°, 20°, 10°

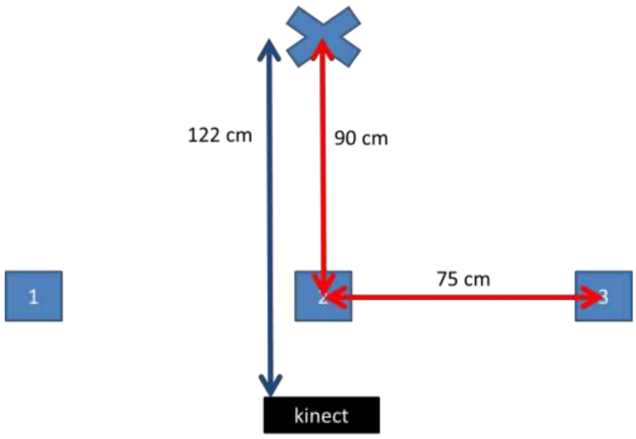
Angle perception is fundamental to prepare children for learning more complex geometrical concepts. Discrimination of angles can be difficult (see the pedagogical section of the Chapter), especially by the youngest children [Mitchelmore and White, 2000]. Moreover, in the context of cross-modal correspondence literature, angle perception is a less investigated geometric feature. In order to provide the means for the employment of other sensory modalities other than vision in teaching angles, it was tested whether cross-modal correspondences were present for angle perception and if that changed with age. To this aim, it was employed a similar methodology to the one described above. In this case, subjects were presented with segments composing apertures of different degrees (angles: 10°, 20°, 40°, 60°, 100°, see Fig. 4.11). The task was the same as described above, subjects were presented with pure tones of different pitch levels (250, 500, 1000, 2000, 5000 Hz) and were asked to touch the monitor to select the aperture that matches the auditory stimulus.

These results showed the same effect observed for size perception. High pitch sounds were associated with small apertures while low pitch sounds were mostly matched with large apertures. This trend was true for all the tested school level. The analysis of variance (ANOVA) showed that there was an effect of the sound pitch level on choosing the angle that corresponded to the sound pitch whereas there was no influence of the school level nor of the interaction of the two factors . These findings indicated that pitch could be used as a non-visual sensory cue to help discriminate angles of different apertures. Future additional analysis may reveal whether there is an effect given by school level and age, unveiling potential differences amongst children.

4.4.2.3 Transformations: triangle completion with sound: audio position information can be useful to determine the starting point of a shape

Previously, the triangle completion task was presented as a measure of children's sensorimotor capability to perceive angles and geometrical shapes during movement without vision, thus using self-generated vestibular and proprioceptive feedback. Nevertheless, external sensory information is often used to update our relative spatial position as we move through the environment. To test whether children were able to integrate internal and external sensory cues during spatial updating, we collected preliminary data with an experiment aimed to disclose such perceptual properties during development. To this aim, IIT UVIP ran the same experiment described in previous section on psychophysics experiments (see Paragraph 4.3.3) by adding an auditory stimulus to be used as a cue for spatial updating. The experiment was conducted on 9 children, 5 frequenting the 5th level of primary school and 4 frequenting the 1st level. This allowed us to collect preliminary data on how such multisensory integration changed at the extremes of our age and school level range of interest. Before each trial, a speaker located behind the participant emitted pink noise sound stimulus and thus indicated the home position. In order to test whether children integrate the auditory cue while moving, we varied the position of the sound source only after the trial started - the sound source was physically moved in one of three possible positions, one that coincided with the home position (no conflict) and other two that could be 75 cm either to the right or left of the home position (see Fig. 4.12). Once the first part of the movement has started, i.e. while walking along the first two legs of the triangle, the sound cue was presented one step after the child has started moving with the experimenter. This was done in order to make the conflict not recognizable by the subject. The auditory cue was turned off when the participant finished the movement along the two legs of the triangle such that the last part of the movement (when the subject has to actually complete the triangle) was done with no sound cue. Results obtained were comparable of subject performance to the simple triangle completion task with no sound cue (see paragraph 4.3.3) with the condition here described will allow us to test whether an auditory landmark can influence path integration. This experiment was important since provided information on how children integrated external auditory cue with internal self-motion cues during movement through space. The influence of such cues was fundamental for the development of weDRAW serious-games.

Figure 4.12: **Illustration of sound cue position.** Numbered blue boxes indicate the three possible positions where the sound source could be located. The blue “X” indicate the home position.



Chapter 5

The weDRAW platform

5.1 Full-body platform and weDRAW architecture

The technology platform of the weDRAW Project consisted of (i) a low-level synchronized movement capturing layer, (ii) a middleware level responsible of configuring the input, mapping, and output modules, of session management, and of session data archiving, and (iii) a weDRAW device layer for generation of auditory, haptic, and visual feedback.

Figure 5.1: HTC Vive system

Full-body interaction encompasses one or more input devices which can capture limbs and full-body movement in an open space (e.g., a room), a collection of software modules to analyse such movements in real-time, a collection of software modules to map movement features on visual and auditory feedback, and output devices which can render such feedback in the space (e.g., environmental sound and lights, large screens, and so on) and possibly on the child's body. Selection of input and output devices and design of the logical architecture of the weDRAW Integration Platform were performed by keeping into account these two ways of interacting with weDRAW technologies.

r

Figure 5.2: Microsoft Kinect V2

5.1.1 Input and output devices

After analysis of the state-of-the-art and of available commercial solution, concerning full-body interaction, both a high-end solution using the HTC VIVE augmented reality device and a low-end solution grounding on an RGBD sensor device (Kinect v.2) were envisaged. Whilst in the following a short description of such devices is provided, output devices consisted mainly of common video projectors and loudspeakers for full-body interaction.

5.1.1.1 HTC Vive

HTC Vive[1] (see Fig. 5.1) is a virtual reality headset. The headset uses "room scale" tracking technology, allowing the user to move in a 3D space and use motion-tracked handheld controllers to interact with the environment. HTC Vive can be used both as an input device for embodied and haptic control systems and as a feedback and output device using the built-in screen and haptic feedback systems, it can also provide motion tracking of users limbs using tracker accessories. In weDRAW, HTC Vive has been mainly used as an input device for full- body interaction for the "CartesianGarden", the serious-game developed by the partner Ignition Factory (IF) and sonified by us with the design that will be presented shortly.

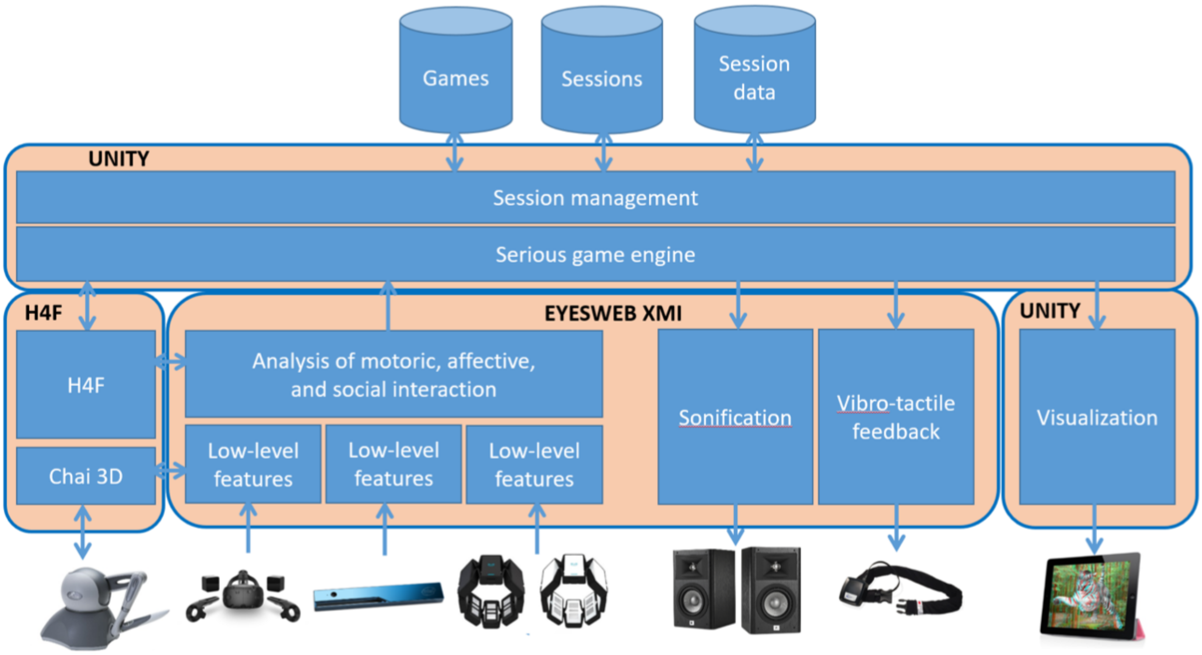
5.1.1.2 Microsoft Kinect V2

Kinect[2] (see Fig. 5.2) provides a natural user interface that allows users to interact intuitively and without any intermediary device, such as a controller. Kinect performs audio, video recordings, and depth computation, i.e., for every pixel the sensor can capture, Kinect measures its distance from the sensor. The Kinect SDK gives access to raw data streams from the depth sensor, color camera sensor, and four-element microphone. Moreover, onboard, the device performs a Skeletal

[1] For more details, please visit: https://www.vive.com/eu/

[2] For more details, please visit: https://developer.microsoft.com/en-us/windows/kinect/hardware

Figure 5.3: The overall weDRAW technological platform



tracking i.e., the capability to track the skeleton image of a person moving within the Kinect field of view. Kinect V2 can track as many as 6 people and 25 skeletal joints per person.

5.1.2 The weDRAW Platform architecture

Fig. 5.3 shows the overall logical architecture of the weDRAW technology Platform, as conceived starting from the use cases of the project, the requirements of the serious games envisaged in the project and the analysis of the available technologies inside and outside the Consortium. As we can see on the top of Fig. 5.3 data repository is represented, including a collection of serious games, and the archive of the sessions carried out by children with the platform, along with the corresponding session data. Following a top-down analysis, immediately below the repository the session management component was in charge of all the operations related to user's sessions with the platform. It managed the creation of a new session, opening or closing an existing one, and association of metadata to sessions. The session management component is directly connected with the serious game engine, which is responsible of running the serious games envisaged in the project

At the bottom left of Fig. 5.3 the input components were sketched. These were grouped in two major stacks. The first one was conceived to support desktop interaction, a part of the project in which our team did not collaborate. It included the interface to the Phantom Omni device[3] and the libraries for real- time analysis of hand and arm movements. The second one was conceived to support full-body interaction. It included the interfaces with the full-body motion capture devices, i.e., HTC VIVE and Kinect v.2, as well as the libraries for real-time analysis of full-body movement. Low-level features were extracted from all the capture signals. These were simple descriptors characterizing the movement being performed (e.g., kinematics, dynamics, shape of trajectories, and so on). The computed low-level features were received as input by the component for analysis of motoric, affective, and social interaction. This component was responsible for a more sophisticated analysis of the user's behaviour including, for example, the affective qualities of her movement (e.g., hesitant, impulsive, prudent, and so on), the expression of emotion related to the learning task (e.g., stress, boredom, interest, engagement, and so on), the social interaction with possible other users (e.g., inter-personal coordination, leader-follower relationships, and so on). For full-body movement capturing and processing, this was mainly performed by adopting the EyesWeb XMI[4] platform developed by our Lab. EyesWeb is a platform for fast prototyping of applications involving synchronized capturing and processing of multimodal data from different sensory channel. In particular, EyesWeb includes a library for analysis of human full-body movement with special reference to expressive movement qualities, a library for analysis of social interaction in small groups with special reference to inter-personal coordination and emergence of leadership, and a library for interactive sonification and active music listening. Moreover, it supports network connection through OSC. As for input devices EyesWeb supports Kinect v.2, Myo, and Notch and other sensors, as well as HTC VIVE. Regarding auditory and visual output, the former can be managed in EyesWeb, the latter either in EyesWeb, or directly in Unity, or using available commercial software. At the bottom right of Fig. 5.3 the output components of the platform were depicted. These consist of three components, one for each envisaged sensory channel, i.e., the auditory, visual, and haptic channel. Consequently, output components were responsible for sonification, visualization, and generation and rendering of vibro-tactile feedback, respectively. The output component were interfaced with the corresponding output devices, i.e., loudspeakers, video projectors, and vibro-tactile devices, respectively. Mapping of input (i.e., the analysed motor behaviour) onto output (i.e., auditory, visual, and/or vibro-tactile feedback) happened at two different levels: (i) in the serious game engine, that is, both computed low-level features and high-level affective and social characterizations of motor behaviour were received, assessed, and a suitable feedback was determined, depending on the current state of the game, the learning targets, and other possible factors, (ii) with a direct connection between input and output components, to enable fast reaction of the system to the user's actions. Finally, the arrows in Fig. 5.3 show how data were communicated between modules. For the sake of simplicity, only the mechanism for feedback generation expecting data to be fully processed and sent to the game engine, where decisions are made, is visualized.

[3] http://www.geomagic.com/en/products/phantom-omni/overview

[4] http://www.infomus.org/

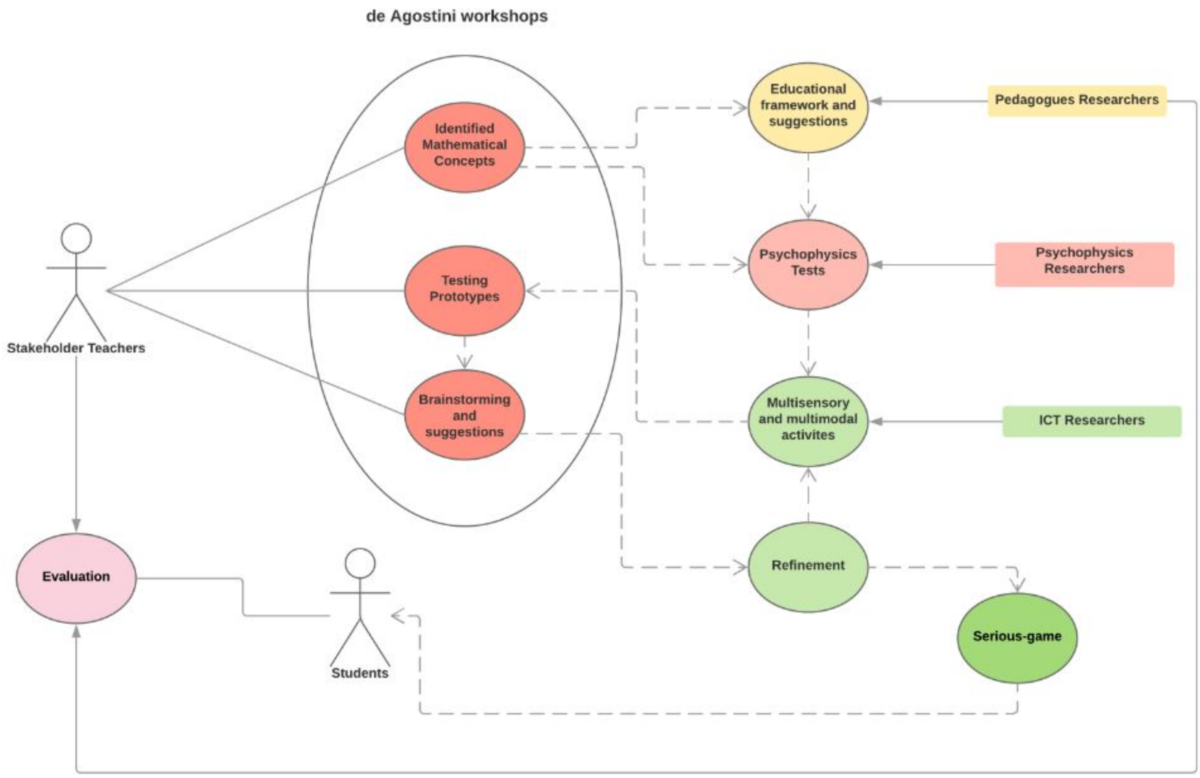
5.2 Game design and prototype development

As Project consortium, we joined the weDRAW definition below of serious-game, as provided by the UCLKL partner. Serious Games in weDRAW were defined as consciously designing games that can function as a vehicle for learning ‘serious’ (i.e. non-game) content and motivate learners in new ways [for Education, 2013]; games and play are an important part of the social and cognitive development of young children [Nicolopoulou, 1993]; serious games initiatives which focus on deeper learning in the context of an enjoyable experience are more likely to succeed in their pedagogical aims [Ritterfeld et al., 2009]. A serious game, then, will not succeed just because it is a game with educational content: it must be engaging and motivating to succeed in deeper educational objectives. Moreover, multisensory experience needs to be embedded in a strong game narrative. Educational games, then, need to be designed with careful attention in particular to:

- contemporary learning theories;
- customisation of task difficulty to the learner’s capabilities;
- metacognitive reflection on the learning taking place;
- consideration of the rich situated interaction among learner, game environment and classroom environment [Young et al., 2012];

weDRAW was a consortium which worked together to define the core ideas underpinning the project. The main idea behind the project was that since multisensory integration develop after 8-10 years of age [Gori et al., 2008], for young children the use of one sensory information can be more useful than the use of multiple. As we see both in Chapter 1 and in previous Paragraphs on weDRAW multisensory studies, Par. 4.3-4.4, other neuroscientific researches showed that different sensory information, such as the audio or the haptic one, can be crucial to understand some specific concepts. Results also suggest that children with visual disability have problems in understanding the auditory perception of space [Gori et al., 2012], [Gori et al., 2013], children with motor disabilities have problems in understanding the visual dimension of objects [Gori et al., 2010], and deaf children have problems in understanding rhythm perception [Reithinger et al., 2003]. For this reason it is important to use all the different sensory channels available and in particular the auditory or haptic sense. This suggests that primary school children would benefit from the use of ICT multisensory technology targeting multiple sensory modalities such as vision auditory and tactile/haptic signals for learning specific concepts (e.g., touch for dimension or rotation and audition for fractions). children. Primary school children are used to play with the games, be engaged by the components narratives, and interact with the game mechanics. One of the objectives of weDRAW was to address the learning needs of typically developed and visually impaired children, and to merge these two classes as much as possible, providing technological solutions that may help overcome barriers to learning. The teachers, classroom assistants and other adults involved in the delivery of learning via the serious games produced, interacted with the game in similar ways to the children. Using traditional methods (keyboards, mice and GUI components) they started and configured the games, selected individual game modules or sections, customised game options for learning classes, groups or individual children, and managed the game sessions. They also encouraged and supported the children in their use of the games. Considering the target of weDRAW games, visual feedback was not the most important one but it was still important to take it into account for two different reasons: first, low-vision children may be able to distinguish shapes or colours. Second, we wanted to exploit how important the visual feedback is in the learning process: would non-blind children be able to perform better with visual feedback, or would audio and haptic feedback be performing enough? Visual feedback was done mostly using video projector. The amount of time a child or group of children should engage in an individual game module, sufficient to achieve the mathematical learning outcome was unknown at the time of development and strictly related to teachers' pedagogical choices. For this reason, all the prototype developed were designed to be playable for few minutes, to be more flexible and adaptable to scholars' needs. Other factors that informed this time value included: (i) whether the game module is being played by 1 or more children and (ii) the complexity of the mathematical concept being covered by the game module. To inform this aspect, we built the game prototypes for 'completion' to be achievable in a short time period – say 10 minutes, considering also teachers' and educators' feedback on how may they managed the weDRAW technologies in the classroom routine. The design of the serious games had a specific focus on the simplicity and ease of configuring the software components, attaching any external devices – Kinect etc. – so that use of the games was encouraged rather than discouraged. The serious games designed needed to be easy for the teacher to understand and use, and for the children to play and engage with. Clear, simple and unambiguous user interfaces and user experience (UX) were developed to be inherent in the overall game design. These was also supported by clear, simple and unambiguous instructions for the teaching team. Serious games and game modules developed as part of the weDRAW project had to cater for children with a full range of visual capability, from full and unimpaired vision, through different types of visual impairment, to near or total blindness. We could not afford to take into consideration the wide range of types of unimpaired vision, but we needed to be mindful of the fact that the serious games had to be inclusive for all children, regardless of visual capability, and that the development of the serious games had to provide core learning for all visual capability while retaining both the fun and engagement to encourage play, and be easily setup and configured by teachers, teaching assistants and other classroom support staff. Having considered these challenges, we developed a set of prototypes, mainly on angles and fractions concepts, plus a couple of sound design for Cartesian plane learning. Our approach exploited multisensory feedback starting from full-body movement interaction. We firstly considered how to covey the same kind of content through different sensory modalities. Furthermore, we based our approach on music, as an effective medium to understand algebraic concepts: music and arithmetic are indeed quite directly associated, based on the suggestion that temporal sequences are at the basis of rhythm and that

Figure 5.4: Flow chart on the interaction design process we implemented. The design and development of the weDRAW technology applied user-centred participatory design methodologies and early prototyping. It indeed required continuous interaction and frequent testing with end users, especially children and teachers



there is an immediate relation between music elements and numerosity (See Paragraph 1.4.2 for an extended review).

The activities described here were presented, in addiction to test protocols dine in schools and visual impaired rehabilitation centre, to children, primary school teachers, families, and the general audience in the framework of some public events organized by the Italian publisher DeAgostini Scuola in Italy. Namely, the activities were available in two booths dedicated to weDRAW technologies at three international science festivals: Festival della Scienza in Bergamo (October 02 – 15, 2017) and at Festival della Scienza in Genova (October 26 – November 5, 2017) and (October 5 - November 4, 2018). In the three events, more than 3000 children could try our activities and in several DeAgostini teachers' workshops across Italy. Based on this collaborative work, we started to design our architecture. Following the discussion with teachers, we developed a set of initial full-body activities, that will be presented in next sections, where different proprioceptive skills and sensory modalities are needed to solve a mathematical problem. Such activities represented our testbed to define a set of requirements that can be summarized as follows:

- The game should be intuitive i.e., it should enable the child to understand the underlying mathematical concept through sensory stimuli, without the need for initial instructions;
- The game should provide an active experience of music content integrated with interactive graphics (e.g., in the form of visual feedback and interactive sonification);
- The game should use meaningful and motivating elements in terms of narrative structures (e.g., sonic elements, music, vocalizations and speech, images and stories) to help both typical and impaired children to contextualize their experience;
- The game should collect and analyse children non-verbal behaviour. This includes supporting automated analysis of motor signals, not only limited to low-level features (e.g., trajectories, speed, distance, and so on), but also extended to (i) movement qualities (e.g., symmetry, energy of movements, rigidity, hesitancy, impulsiveness, and further more), and to (ii) social movement qualities of a group of children in multi-player activities, such as backward learning, synchronization and entrainment.
- The game should be flexible and scalable to different contexts such as, for example, classroom activities involving two or more children, rooms for specific activities for impaired children, dissemination events such as science festivals and more.

The platform workstation consisted of a Kinect, connected to a PC, using EyesWeb XMI [Camurri et al., 2016], [Volpe et al., 2016].

The platform built was used to:

- Design multilevel sequences of games suitable for each child, according to the learning target and the individual needs;
- Store personal information and movement data on every use of the system;
- Perform analysis and compute evaluation, based on movement features and scores, on child performance and progression.

5.2.1 Understanding fractions through the body: the Body Fraction Game

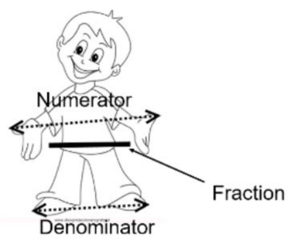
One of the major difficulties we faced in designing activities for fraction learning, was to define a clear representation of the concept itself, despite cognitive, learning or sensory impairment, and involving multiple different sensory modalities. To design this activity, we considered all the interdisciplinary suggestions from literature. The serious-game we present was designed as follows:

- The interaction was based on child's limbs movement tracked with Kinect V2;
- The game was designed to be played by one child at time and the feedback provided is an audio/video feedback;
- The game UI was designed using Unity and connected to EyesWeb XMI.

The game has the following features: (i) it has a multilevel structure to be suitable for each child, according to learning target and individual needs; (ii) during a gaming session, movement data are stored in order to perform analysis and compute evaluation on child's performance and progression.

Our pedagogical target was to teach children the concept of fractions in general, i.e. the relation between the fraction and the unit and between different types of fractions. We designed and proposed a game based on the metaphor of the body as a fraction representation.

Figure 5.5: Sketch of prototyped full-body interaction



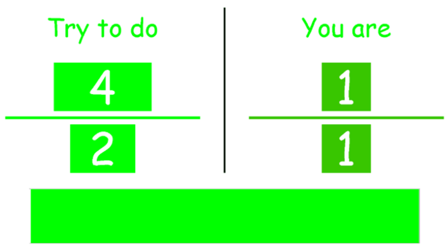
When the game started, the child could explore different types of fraction moving her/his limbs to create the numerator (upper limbs) and denominator (lower ones) of the fraction and listening to the different rhythmic pattern produced (see Fig.5.5). The system detected arms and legs aperture and maps it in the interval [1, 10]. The game interface asked the player to reproduce a randomly generated fraction, displaying a numerator and a denominator, e.g., 3/10. Each fraction produced a different rhythmic pattern. The child tried to reproduce a fraction by moving her/his arms and legs (widening and resting) guided by the auditory feedback, until the desired required fraction is reached. The auditory feedback we proposed, was very simple: a "beep" with a default pitch, which was repeated along time at a given tempo. Tempo was measured in bpm (beats per minute). When arms and legs apertures were equivalent, the system recognized a unit (e.g., 3 /3). Then, the typical heart beat rate, i.e., 60 bpm, was mapped into the sound model. If the numerator was larger than the denominator, two changes were applied to the auditory feedback: (i) tempo is increased; (ii) pitch is increased.

The maximum tempo and the highest pitch were reached if the player kept a very small amplitude in the arms equivalent to number 1, and a very large amplitude in the legs, equivalent to 10. In addition to the rhythmic feedback, the system presents a visual representation of the fraction and an audio/visual comment helpful to link the fraction the child made with the body and experienced as an audiovisual stimulation, to its specific symbolic arithmetic representation (e.g., 1/2, 1/4, and so on). This last point is particularly relevant from an educational point of view, since as teachers and educators explained us, one of the difficulties in understanding fractions for children is to understand the relation between the quantities itself and how that quantity is symbolically represented in arithmetic (see Chapter 4 for more information).

5.2.2 Understanding Fractions through body rhythms: the FractionMusic Game

Another interesting suggestion we received from previous research and from discussion with teachers, concerns the special relation between music and mathematics related to rhythmic pattern. By working on rhythmic music elements, a quite natural learning can be obtained on topics such as the fractions, for example.

Figure 5.6: First prototype of the Rhythm Fraction Game UI



Starting from existing work related to body percussion as a way to convey mathematical concepts[5], we worked on a prototype activity in which we use rhythm and the ability to reproduce it to help children to understand fractions firstly as rhythmic elements and then with their mathematical symbolic representation. Analogously to the version based on the distances between limbs (see previous section), the rhythmic version also aims at teaching children the concept of fraction, but using a different paradigm, i.e., it represents a fraction as a time interval divided by rhythm in smaller intervals. The duration of each bar (or measure) represented the unit (e.g., 1/1, 2/2, 3/3, and so on). The child was asked to divide the duration of the bar into equal parts whose number varies from level to level. In order to accomplish this task he or she had to beat his or her hands. In order to guide the child through the game, the system displayed a line representing the unit. The levels varied in difficulty, changing both the duration of the bar and the fraction to be represented. Literature showed that rhythmic movements were easy for children to understand and that the connection with the symbolic representation of mathematics was quite natural, even in context not strictly related to mathematical teaching. This type of embodied learning helped children with a preference for temporal knowledge to gain a motor knowledge of the concepts. Furthermore, it could be used with visually impaired children for helping them to develop mathematical knowledge leveraging on the knowledge they have of the temporal domain.

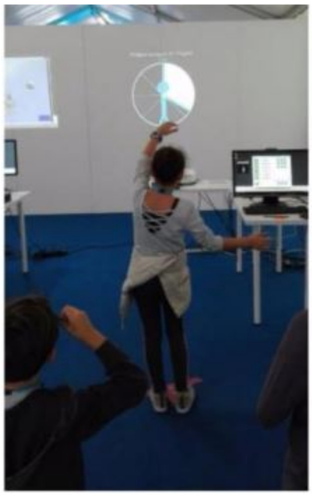
[5]A educational method that based on the use of body percussion for primary schools children can be found at: https://crosspulse.com/rhythm-of-math.html

5.2.3 Understanding angles through the body: the Angle Game

As teachers suggested us, a difficult concept to master related to angles is rotational invariance, i.e., learning that the same angle (e.g., 90°) does not change when it is rotated, even if it is not presented with the orientation children are used to.

Furthermore, for visual impaired children, the relevant problems are indeed more specific and related to the specific spatial domain of the concept, which for these children is very difficult to master. To address such special needs, we proposed an activity consisting in reproducing a specific angle or type of angle (e.g., an acute angle) using the upper limbs, with the help of an acoustic/visual feedback (see Fig. 5.7).

Figure 5.7: First prototype of the visual feedback for the Angle Game



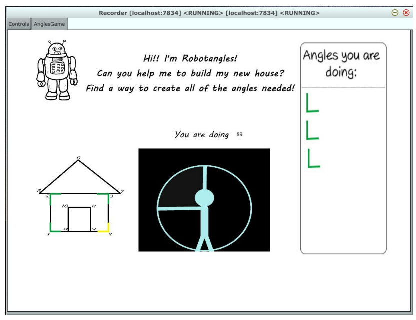
The child was free to move the upper limbs in the space, as her/his head was considered the origin of the angle. While the child moved her/his arms, s/he could hear a musical scale that covered the full range of angle amplitude. If the child changed the opening of the angle s/he was creating with the upper limbs, the note in the scale – which a string instrument plays – changed according to the movement. A long distance between the arms (i.e., a big angle) corresponded to a low-pitch note, whereas a short distance (i.e., a small angle) between the arms corresponded to a high-pitch note. If the child was able to keep the same angle while rotating it, s/he heard the same note, with no changes in the sound feedback, to show that the angle did not change during the rotation. A visual feedback showing the angle the child was creating can also be presented to the child together or without the acoustic feedback. The activity was organized on different levels, from very beginner's knowledge (e.g., acute angles vs. obtuse angles), to more specific knowledge (e.g., 60° or 90° or 180°), to finally master the same angle in the space, rotating it and observing that if the rotation was right, the feedback remained stable. The nature of the acoustic feedback was based on psychophysical results showing that from very early age, children have the knowledge of multisensory association between big shapes and angles with low pitch and little shapes and angles with high pitch. Following this evidence, we sonified upper limbs movement by mapping angles to sound pitch so that acute angles corresponded to a high pitch and obtuse angles corresponded to a low pitch. In the more complex levels, where children have to rotate an angle in the space, sound led the rotation and helped the child to understand how the same angle can be represented in the space using different orientations.

5.2.4 Final version of the weDRAW activities

5.2.4.1 The AngleShapes Game

During the months between July and September 2018, we reworked the design of the proposed activities in a circular iterative design process mainly in collaboration with the UCLKL and UCLIC teams from UCL, London. After this revision, the activities were re-designed to be more engaging and coherent with the overall experience of the weDRAW serious games, made by the two industrial partners. We referenced in footnote[6] a video of the final version of the activities here presented. Considering the Angle Game, this game was redefined by linking the activity of angles creation based on arms' aperture tracking, with the shape creation activity (see Fig. 5.8). We then called the final game prototype the AngleShapes Game. In this way, the child engaged in the game can create, using her arms, a series of angles needed to create a complex shape (e.g., in the example that was developed, the child can finally create a house).

Figure 5.8: The AngleShapes Game UI

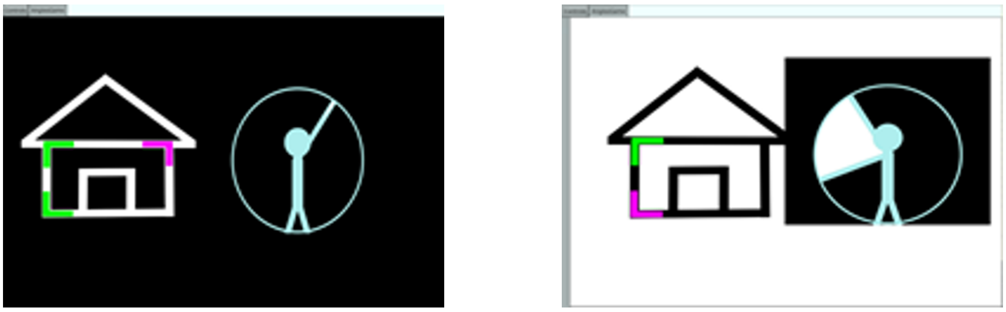


The sonification strategy did not change and the psychophysics association between sound pitch and angle size was used as in the initial prototype. For the user interface, we provided the interface shown in the Fig. 5.10, to help the child to associate his/her arms' movement with the angle created. The list on the right of the interface showed to the child the number and type of angles s/he already did, to help him/her to count the angles created. The house representation on the right is used to track the progress of the child in building the house, relating in this way his/her proprioceptive perception of the created angles with the representation of the house as a 2D shape. The AngleShapes Game was also redesigned to be autonomously used by low-vision children, in two versions both at high contrast, as shown in the picture below.

5.2.4.2 The MusicFractions Game

Concerning the FractionGame prototype, we followed the same interaction procedure used to redesign the AngleGame. For this activity, we put together both the metaphors we explored

[6]Please, here you can find a video of the activities presented in this section: https://www.youtube.com/watch?v=xBTIZwM7S10



(a) High-contrast black Low-vision UI (b) High-contrast white Low-vision UI

Figure 5.9: High-contrast simplified version for low-vision students of the AngleShapes Game

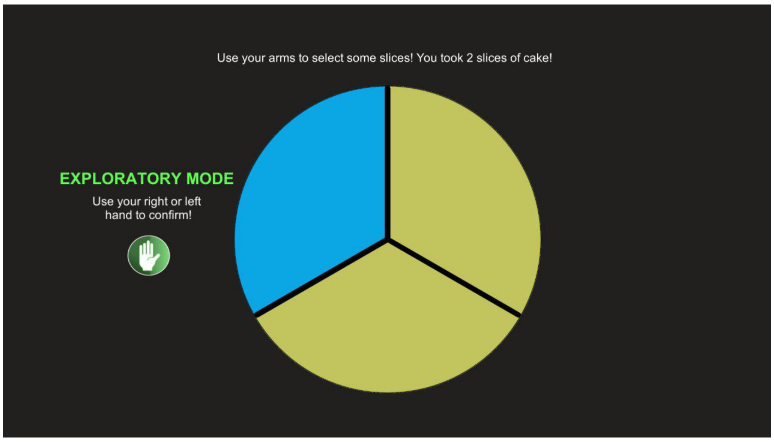
Figure 5.10: A sketch of the sonification used for the MusicFraction Game final prototype

with the previous prototype: in particular, we maintained the idea of using the child's body as a metaphor of the fraction, let him/her to use the arms to define the numerator of the fraction and the legs to define the denominator. We also maintained the rhythmic idea of fraction, using a drum based sonification, to highlight the meter division of the denominator, and the beat determined by the numerator, as in the figure below.

We also redefined the interaction with the interface, defining different game modes:

- Exploratory mode: the child is free to explore the game, opening the legs to divide a circle in how many pieces s/he wants. To confirm the subdivision (i.e., the denominator of the fraction), the child has to "press" the button presented in the interface with his/her left hand, waiting for the button to become green. After this first step, s/he can decide how many pieces of the circle to select, using his/her arms aperture. The slices will change

Figure 5.11: The MusicFraction Game UI



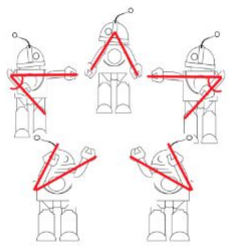
color. In this way, the child can have both the proprioceptive feedback and the visual-auditory one to guide him/her in understanding the fraction made.

- Play mode: the teacher defines one target fraction, through his/her dedicated interface and the child has to follow the same steps presented above, but to create the given fraction. If the child, during the activity, moves his/her legs, the game automatically restarts from step 1. In the picture, you can see an example of user interface for the MusicFraction Game.

5.2.5 Social activities design

Finally, two proof-of-concepts for social game were designed and presented at the IEEE GEM conference in 2018 [Volta et al., 2018a].

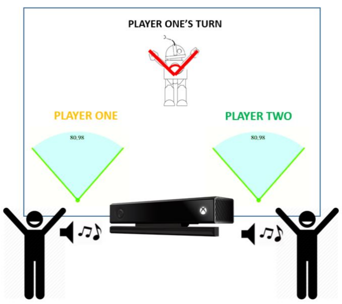
Figure 5.12: Example of a more complex figure to be created with children's arms movement



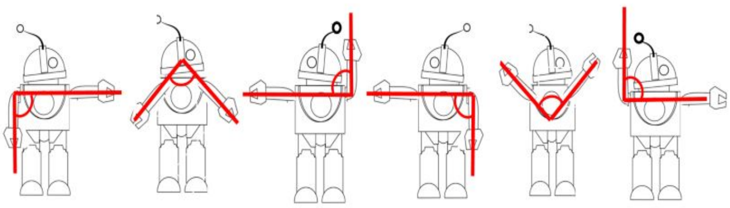
As for the other games developed, also in this case our design process started with several workshops with primary school teachers, organized by the Italian books editor De Agostini in different Italian cities. A relevant suggestion, highlighted by teachers, was to provide students with a way to explore the concept of angles, by involving full-body movement and social collaboration. In this way, both typical and impaired children would be able to play together, doing the same kind of experience each one by using the sensory modality that fits best for her. We proposed a game design to answer this particular request. The prototype mapped body movements onto video and audio, to let the children experience multisensory learning, understanding the connection between proprioceptive and sensory feedback, creating and understanding the mentioned association between arithmetic and music. The game was designed to be used the same settings of the other activities developed. The serious game we presented was designed with the same child's interaction of the AngleShapes Game, meaning the tracking of the upper limbs movements, but with two children playing. The feedback was designed to be provided both with visual and auditory modalities. The game has the following features:

- A multilevel structure to be suitable for each child, according to individual needs and preferences;
- During the game, movement data are stored in order to perform motoric and affective analysis on child's behaviour, performance and progression.

At the beginning of the game, the system asked two players to move their arms so that they represented a specific angle in the space (see Fig. 5.13a). Players' movements were tracked and used to compute which angle was represented by both of them.



(a) A simplified representation of game interaction.



(b) Sequence of angles presented by the system.

Figure 5.13: Sketches on two players Angles Game

Once the players were able to reproduce correctly the required angle, a new angle was proposed (the full sequence of angles is shown in Figure 5.13b). To allow the players to adjust their movements, a twofold feedback was provided: (i) a visual feedback (i.e., two lines drawing the angle created by the arms of each player); (ii) an auditory feedback (i.e., a different sound for each angle). A sound model mapped each angle to a different sound, starting from a reference sound (associated to 0 degrees). The model was designed to modify the sound making it progressively deeper as the amplitude of the angle increased until it reached the deeper sounds, reserved for the angles up to the flat angle (180 degrees). To avoid sounds overlapping, two specific timbres had to be chose by the two players and the beginning of the game session. After such initial phase, a virtual character (shown at the top of Fig. 5.13a) asked the players to help him accomplishing a task, e.g. building a house. In such a way, the players needed to collaborate to complete the task, as shown in Fig. 5.12.

For example, in Fig. 5.14 players had to represent two angles to create: (i) the house floor (ii) the ceiling, and finally (iii) the roof. Once a pair of angles is correctly reproduced, a new pair of angles was highlighted. The level of complexity of the mathematical concepts behind the game tasks was modulated by using different game levels (see, for example, Fig. 5.12), allowing the same platform to be suitable for both typically-developed primary school children and visually impaired ones. At the end of the game session, the platform proposed to the children a generalization of what they just did, to helped them understand the relation between the embodied experience of the game and the notion of angles as they learn it in school. The second proof of concepts was dedicated to fraction learning. For this activity, we deeply used as a metaphor the music performance technique of body percussion. The core idea was to let children create poly-rhythms to explore complex fraction properties (e.g., minimum common multiple), creating each one a pattern, and interacting with the technology.

Figure 5.14: In the second part of the game, a virtual character asked the children to collaborate together to create complex figures e.g., a house.

A sketch of the idea of creating a sequence of movement, as it is in codifying actual body percussion, is presented below (see Fig. 5.6). The child has a reference movement (e.g., the clap), to know the beginning of each movement and rhythmic pattern and a sequence of movements used to represent a rhythmic pattern of 8 beats. This sequence can be used to create a complex fraction or a rhythmic sequence that, combined with the one of the other child, can intuitively (from a musical perspective) show the minimum common multiple.

5.2.6 Cartesian Plane: mapping sounds

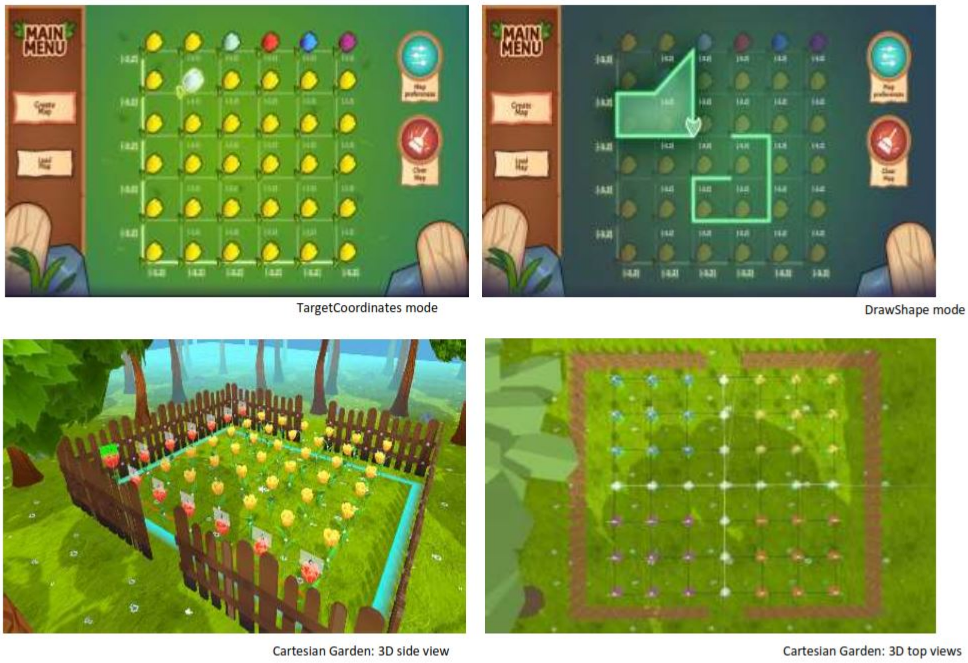
For the Cartesian plane concept, that, as we see at the beginning of the Chapter (See Par. 4.2), was a difficult concept for students to master, one of the industry of the weDRAW consortium, Ignition Factory, provided a virtual reality game, using HTC Vive hardware and EyesWeb XMI. The game was designed to be a garden where the child could freely move and map the space while exploiting it. The haptic feedback was provided by Ignition Factory itself, using a grid of

Figure 5.15: Sketch on proof-of-concept of rhythmic fraction game for two users: the sequence of movement is codifying following the body percussion vocabulary. The two children play in parallel with the system to create poly-rhythms and fractions

box colliders. A box colliders is a 3D virtual box used to detect collision with any other virtual game object. In the Cartesian Garden game, audio and haptic feedback events could be triggered when the controller hold by the player was in contact with a box collider. Those box colliders were placed on the coordinate lines and pillar-shaped box colliders were placed on the coordinate points (see. Fig. 5.17a).

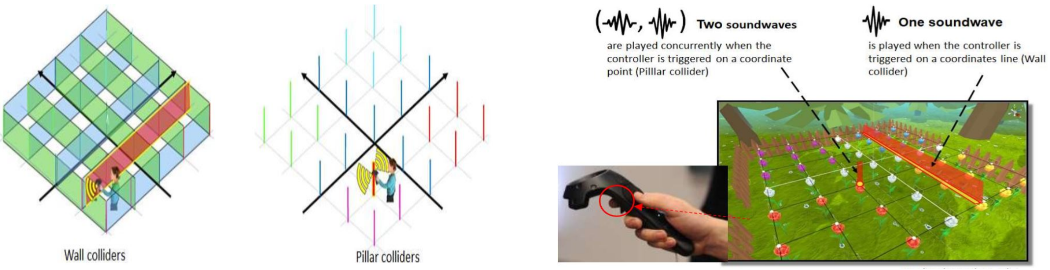
Walls and pickets caught collision with the HTC Vive controllers who responded by vibrating, giving the impression of touching it and allowing the player to understand that s/he was touching or crossing a line (wall) or a dot/node (picket). For what it concerned the audio feedback, we provided the sound design for the game, considering visual impairment in defining the psychoacoustic stimuli to be used. The sound was trigger when the player was in contact with a flower, so that s/he could get the coordinates of that flower where s/he is. The coordinates were given to the player in the form of audio feedback (we linked a sound wave to each wall-shaped collider). The silence time between two sound waves was configurable. We developed two different designs, one with musical notes sounds and one with a landscape sound narrative. Those two designed were intended to let the CartesianGarden be suitable for children of different aged, starting from the younger (even prior than 6 y.o.) that had no studied yet Cartesian Plane at school. A third design was finally done, to let the game be playable by blind children that could not use visual feedback from the virtual reality system.

Figure 5.16: CartesianGarden Game visualization



5.2.6.1 Musical notes mode

In this mode, an instrument was assigned to the abscissa (x-axis) and another instrument to the ordinates (y-axis). The magnitude corresponded to the number of notes played. We use: (i) high pitch for positive integers; (ii) low pitch for negative integers and (iii) specific sound for zero magnitude.



(a) Box collider sketches (b) CartesianGarden Game Music notes mode

Figure 5.17: Box collider and music notes mode sketches

In a scenario example (see Fig. 5.19), a teacher may ask the child to pick up the flower at coordinates (1,-2); the child then moves on the grid and when the controller starts to vibrate softly, the player understands that she or he is touching a line/wall. At that point, pressing the

Figure 5.18: CartesianGame Music sounds mapped into the virtual scene

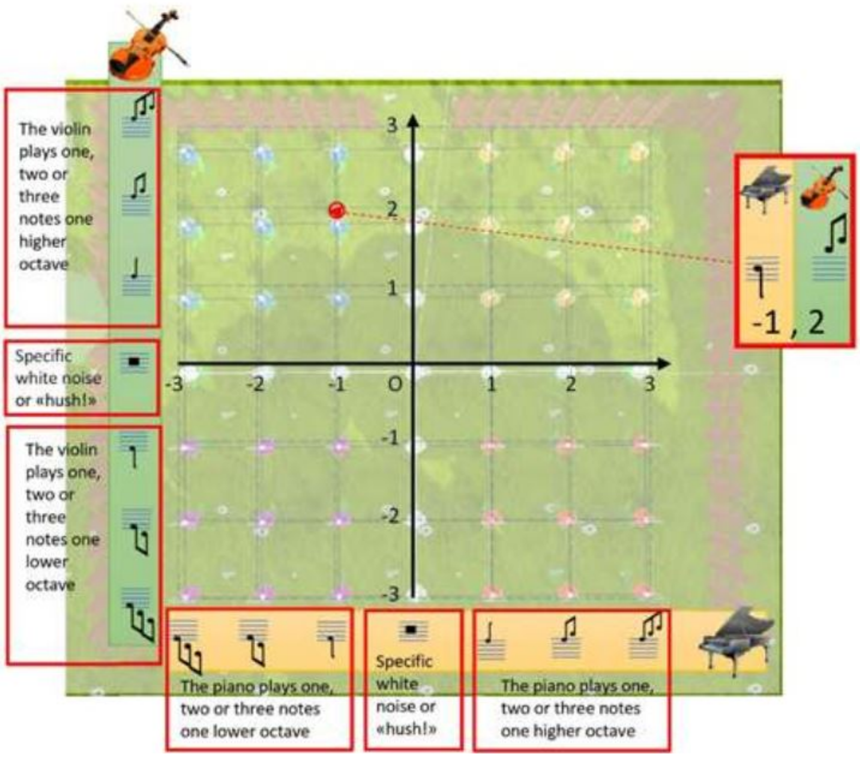
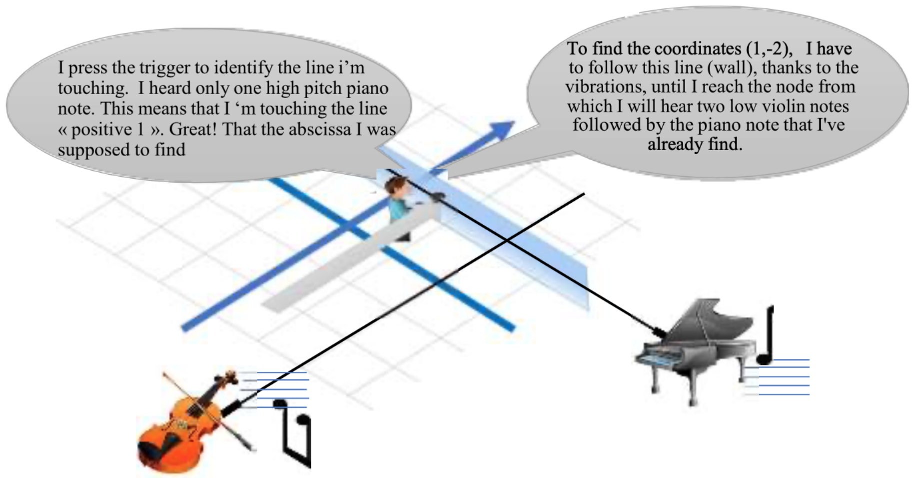


Figure 5.19: CartesianGame Music sounds mapped into the virtual scene



trigger, the child can *listen* the coordinates of that line, checking whether she or he reaches the right coordinates or not.

The second sound designed proposed, was based on a ”narrative” sound idea, suitable for children from 6 years old and younger. That sound modality was an animal voices sound mode, in which there was no notion of magnitude. An animal voice sound was assigned to each line coordinate. The educational goal here for the player was to understand that a coordinate point corresponded to the intersection of 2 coordinate lines. It was intended as a preliminary work to be done with children that does not have prior knowledge of the Cartesian plane concept.

The third and last sound design made for the Cartesian Garden serious-game, was developed to make the game be suitable for blind children. For low-vision children, the author works through a series of usability questionnaires in which the author involved 5 low-vision children and 3 orthoptychs from the Chiossone Institute for visual impairments. This usability work, that will be presented in the last section of the Chapter informed a particular high-contrast version of the CartesianGarden developed by Ignition Factory. For blind users, therefore, a specific sound mode had to be provided, in order to let them freely play with the game without external mediation. The first objective of this modality was then to develop a way that let blind users to walk confident in a straight line along the coordinate lines.

Educators and rehabilitation specialists of blind children asked us to use only the first quadrant, starting their exploration always from the origin (0,0) where the teacher led the child, and the controller fixed on child’s body. We created a ”sound corridor” system that guided the user through the Cartesian plane and helped him/her to walk in a straight line. A continuous sound was played in loop, while the controller was within the limits of the collider. When the player came out of the collider the sound stopped, in the same way as for vibrations; finally, when the player reached a “new magnitude” she or he heard a bell ringing only once (no matter the

Figure 5.20: CartesianGame Animal voices sounds design

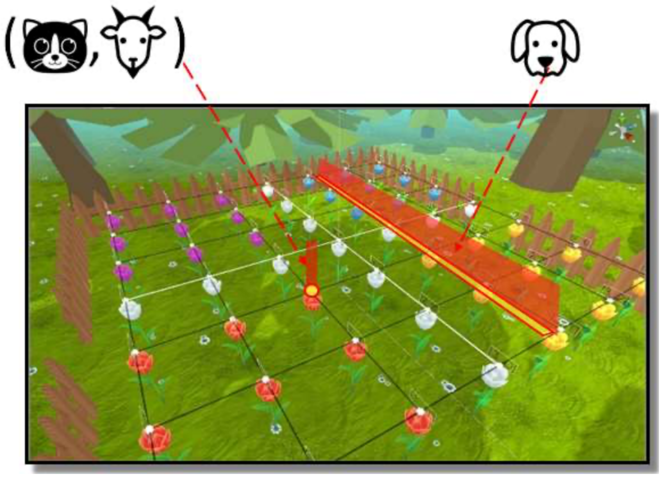


Figure 5.21: Corridor sound model for blind user game modality

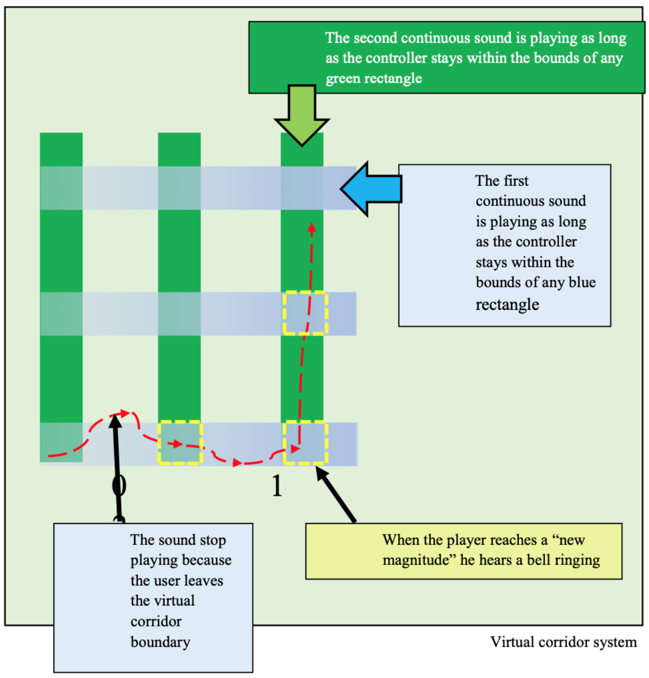
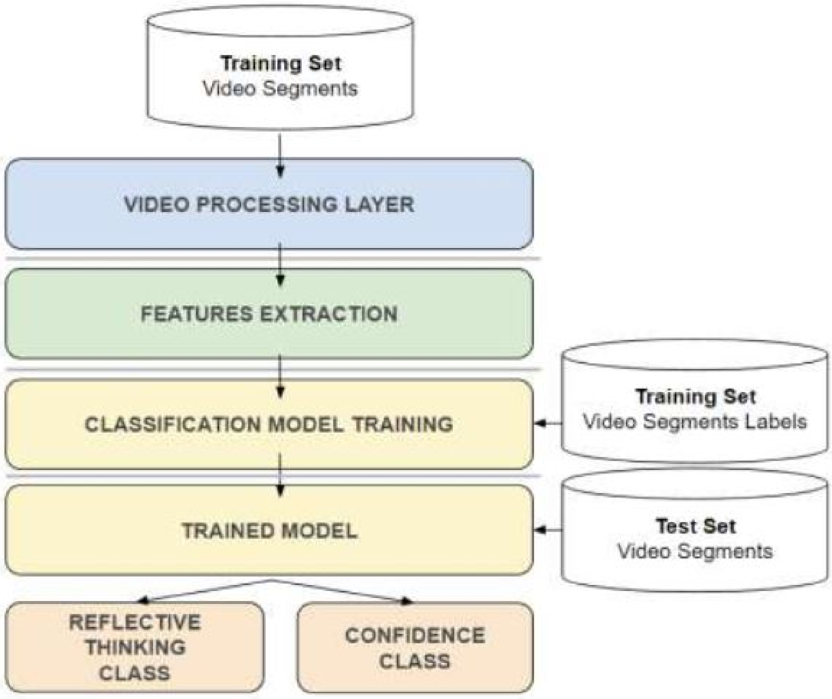


Figure 5.22: Offline video analysis, feature extraction and classification procedure



magnitude), so that the player had to count the number of times the bell rang, mapping in this way his/her path into the Cartesian quadrant and reaching the targeting point. Actively engaging the child in counting how many bells heard was an essential element for the functioning of the game modality, since in this way the agency of the child helped him/her in maintaining the attention on mapping the environment to find the target.

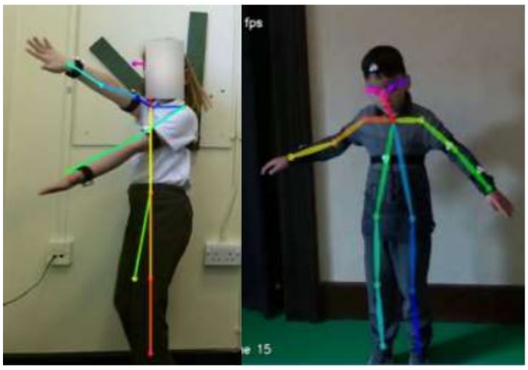
5.3 Automated analysis of cognitive states

We explored the use of computer vision techniques for automatic detection based on video data. In these investigations, we followed the pipeline illustrated in Fig. 5.22, by extracting positional data from the video, before then extracting motion features from these data. These features were then used for training and testing classification models.

We investigated the potential use of OpenPose for analysis of video recordings. OpenPose[7] is an open-source software tool for multi-person pose estimation originally developed at Carnegie

[7] github.com/CMUPerceptual-Computing-Lab/openpose

Figure 5.23: Illustration of the output provided by openpose on an image file



Mellon University. It makes use of deep neural networks and is built with the Caffe[8] deep learning framework. It can be used to reconstruct body pose from images and videos. Fig. 5.23 illustrates a basic use of the software on image files. Pre-recorded video files and webcam video streams can be processed in a similar fashion.

A total of 18 anatomical joints are tracked by OpenPose. We used a subset of these in our feature discarding joints that we expected to have strongly noisy data. The features were computed on 6 joints: the front head, left and right elbow and knee, and C7 cervical vertebrae. 17 features were computed in total. The details of these features are given in Table 5.1.

No	**Name**	**Description**	**Joints**
1-2	Left and Right Arm Velocity	Speed of an upper limb computed frame-by-frame and filtered using low-pass filter	Left and Right elbow
3	Head Velocity	Speed of a head computed frame-by-frame and filtered using low-pass filter	Front head
4-5	Right Arm Energy and Left Arm Energy	Kinetic energy of an upper limb computed frame-by-frame and filtered using low-pass filter	Left and right elbow
6	Head Energy	Kinetic energy of a head computed frame-by-frame and filtered using low-pass filter	Left and right elbow
7-8	Right Arm Variance and Left Arm Variance	Variance of an upper limb position computed on the sliding window of 50 frames	Left and right elbow

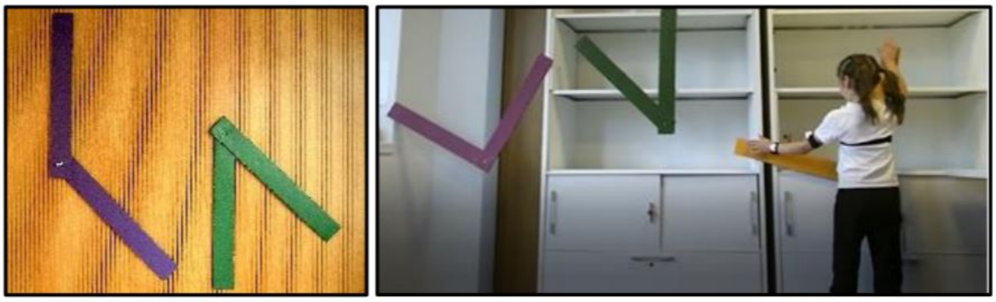
[8]caffe.berkeleyvision.org

9	Head Side Leaning	Difference between the x coordinate of the head and the mean of the x coordinates of the upper limbs	Left and right elbow and Frontal head
10-11	Right Arm Stability and Left Arm Stability	Absolute difference between the raw and the filtered positions of an upper body limb	Left and right elbow
12	Energy Pauses Index	Number of time intervals in which the summed kinetic energy of upper limbs is lower than a given threshold	Left and right elbow
13	Arms Fluidity Index	Average movement fluidity computed on the both upper limbs	Left and right elbow
14	Closure Area	Area of the rectangle drawn by the position of the head and upper body limbs (inset box). This feature indicates the amount of space occupied by the body	Left and right elbow and frontal head
15	Head Rocking Regularity	If peaks of the head movement appears (nearly) regularly on the 50 frames sliding window the output is 1; otherwise is 0	Frontal head
16	Body Symmetry	Index is the filtered sum of: • the sum of absolute differences between x (resp. Y) coordinates of the upper body limbs and C7‘ x (resp. y) coordinate; • the absolute difference of the head x coordinate and C7 x coordinate	Left and right elbow, C7 and Frontal head
17	Foot Symmetry	Difference of the distances between x coordinates of a lower body limb and C7 x coordinate	C7, Left and right knee

Table 5.1: List of Body Movement Features Extracted from The weDraw Dataset

In [Olugbade et al., 2018b], we described the first work done to automatically labelled cognitive

Figure 5.24: (a) Left - 135° and 45° Angle-Arms attached to the wall, for the Bodily Angles Sums task; (b) Right - A 6-years old child gauging the size of a 90° Angle-Arms using her arms, in an adapted version of the same tasks, with two additional AngleArms: another 90° and a 45°



states from children's non-verbal communication in learning tasks. As we stated in the paper, the weDraw-1 Movement Dataset was collected in London with typically developing primary school children, in two main settings. In Setting A, we used a room within the university and there was one of the child's parents (and sometimes one or two siblings) present in addition to the child and two researchers, one of whom interacted with the child and acted as the instructor. In setting B we used a (smaller) room within the primary school that the child was attending at the time of the study, with only the two researchers and the child present. All the data were collected in the UK, by our consortium partner of UCL UCLIC, and all the children were studying in the UK at that time, in school years between Year 2 and Year 7 with an average of 4.38 (standard deviation of 1.47). The data was collected from a total of 26 children (14 children in Setting A and 12 in Setting B) between 6 and 11 years old with mean age of 8.69 years and standard deviation of 1.19. The dataset comprises 120 sequences (64 in Setting A and 56 in Setting B) of video with corresponding three-dimensional full-body positional data, each of a child performing a single task or multiple related tasks. The sequences are from 24 children whose parental consent permitted us to get their videos annotated. The longest sequence is 537.23 seconds in length (median = 117.94, interquartile range = 154.18). Both the video and positional data were captured using the Microsoft Kinect v2 sensor. To support the core idea of the weDraw project, the tasks were designed to create real learning experiences and several pilot tests were carried out to iteratively improve the interaction and exploration experience of the child participants. The problems in the tasks were informed by pedagogical studies partially published in [Price et al., 2017], with a focus on angles, symmetry, and shape reflection. There were five main types of task (these tasks were performed with repetitions, usually completed in the given order):

- Forming Angles: in this task, the child explored static representation of given angles using their arms (e.g. Fig. 5.7) and received automatic visual feedback (on a screen) based on the first prototype of the AngleShapes Game.

- Bodily Angles Sums and Differences: Here, the child was given a pair of angles, each represented by a three-dimensional object (named 'Angle-Arms') depicting the rays of an angle, e.g. 135° and 45° in Fig. 5.24. The Angle-Arms were attached to a wall and the task for the child was to represent with his/her own arms the angle resulting from the sum (or difference) of the given angles. One strategy to solve the problem would require the child to first place his/her arms against the rays of the first Angle-Arms to represent its angle. The child would then need to go to the second AngleArms and align one of his/her arms against one of the AngleArms's rays, keeping the angle representation of the first Angle-Arms. After this, the child would have to sweep his/her aligned arm (keeping the other still) towards the second ray of this second Angle-Arms. To reduce the complexity of the problem, the Angle-Arms were arranged such that a ray of the first had the same orientation as a ray of the second (see example in Fig. 5.24). The child was encouraged to use his/her body to explore the angles in finding the solution.

- Rotating in Angles: Each child was asked to represent the sum/difference obtained in Bodily Angles Sums and Differences tasks as full-body rotation. Lines on the floor dividing an imaginary circle around the child's feet into octants (see Fig. 5.24) were used to provide visual guide for this task.

- Finding Symmetry: Here, the child was seated in front of a table and asked to choose from a set of large cardboard shape cut-outs (diameter between 35 and 45.5cm) and show the lines of symmetry of the shape, if any. The younger children were usually first asked to explore the basic shape properties: the number of edges and vertices, and the name.

- Making Shape Reflections: In this case, the child was given a duplicate of the chosen shape and asked to arrange the two cutouts such that one was a reflection of the other, with as many reflection configurations as possible. A line taped across the table (see Fig. 5.24) was used to simulate a mirror.

The data were annotated by two of UCL UCLIC researchers (R1 and R2) with experience working with children and also present during the data collection (and so familiar with the tasks given to the children) independently labelled the data. In [Olugbade et al., 2018b] all the annotation procedure is described in details, but raters reviewed their annotations several times and after consulting with one another. To understand how much the raters agreed on the occurrence of reflective thinking, UCL UCLIC team computed a two-way mixed model, absolute agreement, average measures intraclass correlation (ICC) on the labels for the Setting A data (249,126 frames). We found ICC = 0.63, which shows good level agreement [Griffin et al., 2015]. During the annotation, for each time period judged as moment of reflective thinking, the raters also provided the cues that they used in recognising reflective thinking at that specific moment. There were 531 cue reports; each report specified a sequence or concurrence of behaviours, e.g. "... she pauses and then makes a thoughtful expression with her lips then looks up and away and then draws the answer with her finger ..." (R2). Based on thematic analysis methods [Braun and Clarke, 2006], these cues were coded, and the codes were refined until all codes were clearly defined and no new themes emerged. Table 5.2 shows the list of codes that emerged, with examples from associated cue report extracts.

Code	**Examples**	**High level code**
Speak to self		Verbal
Smile		Facial, Mouth
Speech delay	"opens mouth as if to speak, but not speaking" (R1); "seeming to utter 'uhm'" (R1)	Facial, Mouth
Doing something to the mouth with the mouth	"pushes lips upwards and release just before response" (R1)	Facial, Mouth
Other mouth expression		Facial, Mouth
Doing something to the eye with the eye	"beginnings of a frown" (R1); "squints eyes" (R1)	Facial
Other facial expression	"hard look" (R2)	Facial
Finger(s) touching head region	"hands clapped over nose" (R1); "scratches head" (R1); "finger to mouth" (R2); "rubs head" (R2)	Body (hand)
Pointing		Body (hand)
Head tilt		Gaze
Looking into space/ground/ceiling		Gaze
Looking at relevant object (e.g. shape cut-outs), while in non-action		Gaze
Other gaze (change)		Gaze
Forward lean		Body (trunk)
Back lean		Body (trunk)
Pause at the start		Body (whole); Time
Pause at the end		Body (whole); Time
Other pause		Body (whole); Time
Problem solving duration		Time
Slow movement		Time; Body (multiple)
Tentative/cautious movement		Body (multiple)
Fidget	"worries mouth" (R1); "swings ... leg, ... stops, ... starts again" (R1)	Body (multiple)

Other gesture/posture		Body (multiple)
Reminds self of problem/question		
Exploration by own movements		Exploration
Exploration by moving object		Exploration
Systematic solving		
Solution implementation or response		
Gesturing while speaking		
Other		

Table 5.2: List of Body Movement Features Extracted from The weDraw Dataset

The majority of these codes highlighted bodily behaviours, facial expressions, gaze and verbal behaviours were also noted. It was interesting that the raters (both of them) used a verbal cue in their judgments even though they were not provided with aural data. Some of the facial cues used by the raters seemed difficult to specify in vernacular, e.g. “hard look” (R1), “thinking face” (R2). However, the majority of the facial codes are related to cues involving the mouth. Perhaps, these mouth behaviours serve to communicate that expected speech is delayed because the speech content is being generated or processed (in reflective thinking). Other facial expressions involved the eye. We theorise that this class of expressions and the gaze behaviour of looking away may function as a means of avoiding visual distractors and focusing attention on internal processes (e.g. recollection [Dewey, 1933]), to solve the problem at hand. The link between thinking and looking away has been previously noted [Loyau and Aubergé, 2006]. Head tilting was another gaze behaviour highlighted by the raters. This expression may, together with the forward/backward lean and exploration by moving the object to be manipulated also noted, have permitted the child a different visual perspective of the problem. Exploration may have additionally served in physically evaluating generated ideas. It is not surprising that tempo (pauses, speed, duration) were used by the raters to recognise reflective thinking since this is the one cue specified in the definition of the construct. What is interesting to note from the cues reported is that reflective thinking seemed to occur both at the start and end of problem solving. Another intriguing note is that certain forms of fidgeting seem to accompany reflective thinking. It is not clear what the function of these behaviours is, but they may be used to fill the pause characteristic of reflective thinking or perhaps to break the stillness of that pause so as to stimulate ideas. Other bodily cues highlighted by the raters involved self-touching of regions on the head. Providing a solution was also a cue used; although the raters had agreed not to rely on this, it was decided that for some tasks, the motion of completing the task and arriving at the final solution could count toward evidence of reflective thinking. The manner in which the solution is described to the instructor, particularly elaborate use of gestures, was also used to infer reflective thinking. UCL

UCLIC team then investigated the modelling of reflective thinking periods using LSTMNNs. In [Olugbade et al., 2018b] the machine learning experiment is deepen, but the results we obtained were promising, having obtained with Vanilla LSTMNN architecture an average F1 score of 0.79

5.3.1 Non-verbal communication for visually impaired children

Considering the first results obtained on the data collected on typically developed children, we then decided to deepen our analysis of non-verbal communication in visually impaired children. The level of visual functioning can greatly influence early child development. A visual disability may therefore lead to developmental delays, especially if an early intervention does not take place [Zambone, 1989]. Developmental delays can have a negative effect on child's participation both in rehabilitation and educational settings and congenital visual impairments may be responsible for delays in motor development, because children cannot observe external cues for movement [Longmuir, 1998]. Furthermore, poorer immediate problem-solving [Cole and Pheng, 1998] and mathematical skills [McDonnall et al., 2012] have been found in children with a visual impairment. Nonverbal behaviours in visually impaired children were expected to be semantically different from those of typically developing ones. For this reason, considering the state-of-the-art, it is relevant to investigate the extent to which visual impairment may affect the development of nonverbal communication patterns and the ability to produce such patterns in various social interactions. In [Doherty-Sneddon, 2003], for example, the authors proposed Gwyneth's classification of children nonverbal communication. This classification model includes a) body gesture, b) eye gaze, c) face expression, d) body posture, and e) nonverbal vocalization. For instance, head orientation, proxemics, and distance from objects (e.g., walls) might have different meaning [Sleeuwenhoek et al., 1995] when displayed by visually impaired individuals. Iverson and Golden-Meadow [Iverson and Goldin-Meadow, 1997], for example, discussed gestures used by congenitally blind children who never saw gestures before nor experienced their communicative functions. Results showed that visually impaired children produced gestures, but not in all situations as it was for sighted and blindfolded ones. The study suggested that gestures provide speaker with functions that are independent by the listeners. For the purpose of our study, we created, with the participation of visually impaired children, the VI-weDraw dataset, which comprises body movement data, captured during mathematical problem-solving specifically designed for the experimental setting. The tasks were based on project premises and included: angles, symmetry and reflection, considered as a type of isometric transformation of shapes. The dataset consists of two synchronized video recordings (frontal and lateral) with corresponding audio data. All the children recruited for the experiment were studying in Genoa and participating in a rehabilitation programme at Chiossone Institute in Genoa at the time of the data collection. We collected the data from 3 blind and 14 low-vision children. The visual acuity of the collective group ranged from no perception of the light to visual acuity of 1/20 from an eye. To understand the level of cognitive impairment, the verbal QI (QIV) and performance QI (QIP) items of the Weschler Intelligence Scale for Children (WISC-IV) [Wechsler, 2003] were used as well as the IAG and the Griffiths-III [Green et al., 2016] tests. Collectively, the cognitive tests showed that 3 of the 17 children had levels of cognitive impairments.

Table 5.3: The list of nonverbal behaviours and their frequencies of appearance in the dataset

Movement quality			**Posture**			**Gesture**		
Id	*Cues*	*%*	*Id*	*Cues*	*%*	*Id*	*Cues*	*%*
1	Focused, direct movement	34.44	10	Gaze down	38.99	21	Exp. of positive emotions (e.g. laughter)	18.33
2	Jerky movement	25.00	11	Tendency to act	25.00	22	Nervous smile or laughter	13.99
3	Hesitating movement	22.78	12	Listening predisposition	25.00	12	Open mouth	10.00
4	Fluid movement	21.11	13	Body as a reference point	20.00	21	Nodding during tasks resolution	7.88
5	Impulsive movement	20.00	14	Gaze contact with the interlocutor	18.33	25	Grabbing clothes	7.88
6	Inhibited movement	17.22	15	Withdraw from action	16.77	26	Rocking	7.88
7	Not goal-oriented movement	16.77	16	Outward-facing gaze	13.99	27	Lips biting	6.77
8	Slow movement	15.00	17	A loss of posture alignment	13.33	28	Deictic gestures	5.00
9	Misalignment		18	A loss of balance		29	Hands kept together	5.00
	of different body planes	11.11		(feet support instability)	13.32	30	Hands hold behind back	4.44
			19	Posture openness	8.89	31	Touching face or mouth	0
			20	Legs are moved while body is still	7.78			

Following the procedure described in [Olugbade et al., 2018b] with typical developing children, the data collection session started with an exploration of static representation of angles using child's arms, whose movements were supported with sonification technology. Sounds were realized using a tonal scale played by strings instruments, and the pitch was mapped to the inner angle (along the vertical plane) between childś arms. It was played in real-time, according to child's arms aperture (see Fig. 5.7). Differently from [Olugbade et al., 2018b], the children were also trained to use proprioceptive and tactile feedback with a flat wall as reference: standing with the back flat against the wall represents 90° angle. The child's arms extended outwards ipsilaterally (and against the wall) to form a 180° angle. By extending one arm outward ipsilaterally and the other contralaterally, the children formed a 0° angle, while extending one arm outward ipsilaterally and the other in the anterior they created a 90° angle. For each child the task was first explained by the instructor, who helped the child's exploration of angles 0°, 90° and 180° as described above. Next, the child was encouraged to explore sounds feedback of her/his own. Finally, the child was asked to represent 45° and 135° angles. In the second task, each child was asked to represent 45° angle (and 135° angle) by rotating his/her body about the vertical axis. This additional proprioceptive and tactile exploration was suggested by the expert in visual disability rehabilitation working in the project. From the video data, we extracted the episodes in which each child attempt to solve the given mathematical problems. We collected in this way 92 episodes with total duration of 12 minutes and 10 seconds. All the episodes were presented to the annotators in random order and without audio information. Although in the literature there are findings suggesting that for learning related tasks, untrained observers can also provide reliable ratings, we decided firstly to ask four experts (co-authors of this paper) to annotate each episode. The group was composed of one rehabilitation specialist from Chiossone Institute for visually impaired people, and three experts in movement analysis from University of Genoa. They used a 1-to-5 Likert scale, with 3 as neutral state, for each of the degrees of confidence and engagement. Each vote was a result of the informal consensus between all the experts. Meanwhile, they also discussed and jointly reported the nonverbal behaviours observed. The same experts proposed a list of 31 observed nonverbal behaviours (see Tab. 5.3). For the second round of annotations, we chose 20 episodes, from the initial set of 92, which in the above mentioned ranking were annotated as representing high/low engagement and high/low confidence.

Table 5.4: Cues associated with high-level of engagement annotation, expressed in frequency above median value.

id	Nonverbal Cues	% Annotation Frequency
1	Focused, direct movement	25.56
11	Tendency to act	18.99
12	Listening predisposition	17.22
4	Fluid movement	16.11
21	Expressing positive emotions (e.g. laughter)	15.00

Table 5.5: Cues associated with high-level of confidence annotation, expressed in frequency above median value.

id	Nonverbal Cues	% Annotation Frequency
1	Focused, direct movement	31.66
11	Tendency to act	21.77
10	Gaze down	21.11
12	Listening predisposition	15.66
5	Impulsive movement	15.00

Three experts (one psychomotrician and two specialists in orientation and mobility for visually impaired patients) and six non-experts neither in visual impairment rehabilitation or child development, annotated all 20 episodes displayed in random order. We decided to perform a second annotation by both experts and non experts to check whether particular patterns are differently recognized by these two groups of annotators, meanwhile we expected the same types of observations by the three experts compared to those of the first round annotation. Annotators were asked to indicate relevant nonverbal behaviours from the list presented in Tab. 5.3, and the perceived level of child engagement and confidence (defined as above), using a 1-to-5 Likert scale. To measure inter-rater agreement on perceived engagement and confidence, we computed two-ways random ICC average agreement [Hallgren, 2012] of the Likert values. Results are 0.5291 for engagement and 0.4662 for confidence. These ICCs are fair [Cicchetti, 1994], especially considering the mixture of expertise of the annotators. For annotation of nonverbal behaviours, we firstly computed the frequency with which each cue was selected as relevant. The nonverbal

Table 5.6: Cues associated with low-level of engagement annotation, expressed in frequency below median value.

id	Nonverbal Cues	% Annotation Frequency
10	Gaze down	19.44
2	Jerky movement	13.89
15	Withdraw from the action	12.78
3	Hesitating movement	10.56
7	Not goal-oriented movement	10.56

Table 5.7: Cues associated with low-level of confidence annotation, expressed in frequency below median value.

id	Nonverbal Cues	% Annotation Frequency
10	Gaze down	19.44
3	Hesitating movement	14.44
2	Jerky movement	13.33
6	Inhibited movement	10.00
7	Not goal-oriented movement	10.00

behaviours: 1) related to gaze direction, 2) movements qualities, such as directness, hesitation, or jerkiness, 3) involvement or retraction from the action, and preparation to listen were frequently chosen as relevant (see Tab. 5.3). To understand the relation between behaviours and the targeted cognitive states, we computed annotation frequency of each behaviour in only videos ranked as ”high engagement” and ”high confidence”. We used the median value for each cognitive state (for engagement equal to 4 and for confidence to 3) as a threshold for separating the episodes rated as ”high” from the ”low” ones. Next, for each behaviour, we computed annotation frequency separately for videos above and below these thresholds. The most frequent behaviour for each state is reported in Tab. 5.4 (high engagement), and Tab. 5.5 (high confidence). As it can be seen, two of them are presented in both states: *focused movement* and *tendency to act*. Thus, they can be indicators of both high engagement and confidence.

We used the same approach to find nonverbal cues of low engagement and confidence. Results in Tab. 5.6 and 5.7 show that low engagement and confidence are mostly expressed with *gaze down*, *jerky* and *hesitating movement*, and *non goal-oriented movement*. Interestingly, cues appear frequently for low levels of both cognitive states (e.g., *gaze down*, or *jerky movement*).

As reported in Tab. 5.5- 5.7, some cues were frequently annotated: 1) for both cognitive states, and 2) in high and low level of the same cognitive state. This might give to the reader the impression that these cues are not discriminative. There are, however, possible explanations for this observation. Regarding case 1), it is possible that such cues, e.g. *focused movement*, are relevant for a cognitive state only if they co-occur with some other behaviours. The sharing of specific cues between low confidence and engagement may also be explained by the fact that the two states often co-occur. Indeed, frequently low level of self-confidence tends to influence engagement and participation in task, especially considering the learning tasks and children's young ages [Eccles, 1999]. Regarding case 2, it is important to notice that some cues are binary (e.g., *appearance of laughter as expression of positive emotion*) whilst others can be considered as continuous cues (e.g., *fluid movement)*). In the case of continuous cues, it might happen that different ”degrees” of such a cue can be associated with different levels of the corresponding cognitive state. This example shows a shortcoming of our annotation schema, as so far we have not used a continuous scale for the annotation of nonverbal behaviours. Thus, future works are needed to address this shortcoming. Considering these preliminary results and the final aim of the work, it is interesting to highlight that cues as gaze, generally considered as one of the fundamental cues in video detection of children and students engagement [Rehg et al., 2013], [Sanghvi et al., 2011], had a relevant role for human annotators also in the context of visually impaired people, especially in the case of low engagement and confidence. Literature on early-social cognitive development, deeply analysed the use of gaze, starting from infancy, as a privileged cue of social interaction and others' attention and intention. When *mutual gaze* occurs, according to theorists [Meltzoff et al., 2007], it is a sign of social engagement and mutual interaction, whilst *following gaze* is considered a sign of understanding others' attention. *Gaze alternation*, in dyadic or group interaction, is used to assess joint attention [Akhtar and Gernsbacher, 2008]. Those cues were largely considered, for example, in understanding engagement in autistic children

[Chorianopoulou et al., 2017]. From the literature, we know that blind children have difficulties in detecting patterns of social interactions, meanwhile sighted people surrounding them may have difficulties assessing where a blind child focuses her attention, since there is neither visual orienting and pointing, and gazing and facial expressions are more neutral [Bigelow, 2003]. For this reason, we may suppose that the absence of such common patterns of joint attention and engagement, led annotators (who were in majority non-experts), to consider still position of gaze, looking down, as a cue of both lack of engagement and self-confidence, as this is how it is perceived by typically developed people in social interaction contexts. To check this hypothesis (and be able to perform a comparative analysis between non-experts and experts), we need, however, to collect more annotation from experts.

5.3.2 Automated classification

5.3.2.1 Features extraction

To check whether it is possible to detect the levels of engagement and confidence from visual data in the context of a single-user task, we performed a series of preliminary experiments with standard machine learning techniques on the dataset collected. Due to the small number of annotated episodes, we subdivided them into smaller segments of fixed duration of 50 frames (corresponding to 1 second), obtaining 758 segments. The labels for each of the two states were not balanced. The same shortcoming was observed after the sub-segmentation. Thus, we decided to regroup the rates:

- Levels 4 and 5 on the engagement scale were regrouped into high level engagement,
- Levels 1 and 2 on the confidence scale were regrouped into low engagement level.
- Level 3 expressed medium level of engagement.

We obtained in this way, 294 segments for high engagement, 176 for medium and 288 for low engagement and 172 segments for high confidence, 317 for medium and 269 segments for low confidence. 15 features were extracted from 2D positional data obtained by applying OpenPose [Cao et al., 2018] to the frontal view recordings. The features were computed on: 1) front head, 2) left and right elbow, 3) left and right knee, 4) Cervical vertebrae (C7). Eight features are low-level kinematic features: *Right* and *Left Arm Position Variances*, *Velocities*, *Kinetic Energies*, *Head Velocity* and *Kinetic Energy*. *Head Side Leaning* is computed as a difference between the x coordinate of the head and the mean of the x coordinates of the upper limbs. The algorithm by [Niewiadomski et al., 2019] is used to compute *Right Arm and Left Arm Stability*, [Alborno et al., 2016] to compute *Arms Fluidity Index* and [Castellano et al., 2009] to compute *Closure Area*. *Body Symmetry Index* is the sum of: 1) the sum of absolute differences between x

(resp. y) coordinates of the upper body limbs and C7, 2) the absolute difference of the head and C7 x coordinates. Finally, *Foot Symmetry* is computed as the distance between the x coordinates of the child lower body limbs and C7 x coordinate. It is important to state that, as a first step, these features cover only a subset of the cues listed in Tab. 5.5-5.7. The kinematic features are low-level components of various expressive qualities (first column of Tab. 5.3), such as fluidity or impulsivity[Camurri et al., 2016]. The remaining features correspond to some observed postures and gestures (second and third column of Tab. 5.3), e.g., loss of balance (19), posture openness (20). Next, three aggregation operators (average, maximum, and minimum) were applied to the values computed on the single observations. Thus, a 45-feature vector was used for each segment. Finally, the data was normalized using the z-normalization method. We then performed a set of machine learning experiments to automatically classify three levels of confidence and engagement (low, medium, high) from video data. We explored a set of traditional machine learning algorithms: Support Vector Machine with polynomial (SVM-poll) and radial basis kernel (SVM-rbf), Random Forest (RF), BayesNet (BN), and Multi-Layer Perceptron (MLP). The experiments were performed using feature reduction and dimensionality reduction techniques:

- PCA - 16 principal components extracted from the data (obtained with a threshold of 95% of variance covered),
- Greedy - 12 features obtained from Greedy Stepwise Search combined with Correlation-based Feature Selection.

We have used a 10-fold validation procedure. The results for the three-class classification task are shown in Table 5.8. All experiments were performed in Weka 3.8 software[9].

Table 5.8: Results (in terms of F-score) for 3-class classification task

	Engagement		**Confidence**	
	PCA	**Greedy**	**PCA**	**Greedy**
SVM-rbf	0.63	0.63	0.6	0.62
SVM-poll	0.61	0.56	0.56	0.57
RF	0.65	0.73	0.61	0.7
MLP	0.57	0.54	0.49	0.5
BayesNet	0.52	0.6	0.47	0.56

As can be seen in Table 5.8, the best results were obtained with Random Forest and Greedy feature reduction. In general, the results are not perfect, but it should be noted that the experiments were performed on noisy 2D data extracted using OpenPose. Some tasks given to the children

[9]https://www.cs.waikato.ac.nz/ml/weka/downloading.html

required the rotation of the whole body, leading to a lack of 2D data during the child's rotation. We believe that the results could improve using 3D positional data (e.g., from Kinect or Notch[10]). Another important shortcoming is that we have not used all relevant cues from Tab. 5.5- 5.7. By implementing the remaining cues we hope in future to improve the classification results.

5.4 Developmental Dyslexia screening and training

weDRAW had a second specific research question to consider: how multisensory technology may help both in screening and training skills in dyslexic children. Scientific evidences suggest that there is an interaction between rhythm and dyslexia. A first part of work was mainly carried out by our partner of IIT UVIP team, discussing with an expert of Dyslexia (Prof. Usha Goswami) about the ideas and limits of the weDRAW approach. Prof.Goswami on the link between dyslexia, rhythm and math gave us a feedback, highlighting that the strength of this link is heavily grounded on "on the temporal load of any math task, and the temporal characteristics of a rhythm task."

> "There is definitely an auditory temporal processing deficit in children with dyslexia, and we know this impairs performance in mental arithmetic tasks with a memory load (via working memory). Given that this auditory temporal processing deficit has affected multisensory integration over the entire developmental trajectory of a given child, by the time they are in school it is difficult to get a 'pure' measure of anything."

And moreover,

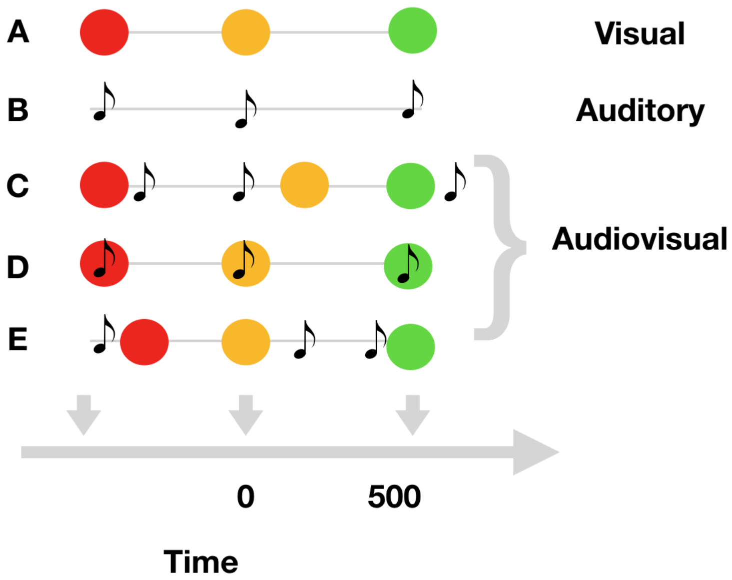
> "[...] for the simpler question of how you can diagnose a rhythm deficit that may indicate dyslexia, the best task would be our auditory rise time discrimination task using the Dinosaur interface. Higher thresholds are typically found in children with dyslexia. The task is not automated, a tester needs to sit with the child while he/she completes the task. However, it only takes around 5 minutes. Another task that we find to be diagnostic is the same-different metrical music task [Goswami, 2011, Huss et al., 2011]. The primary cause of dyslexia in my judgement is impaired auditory processing, specifically of slowly-changing amplitude modulation cues, and most specifically, amplitude envelope onsets (rise times, [Goswami et al., 2002]). Amplitude modulation (AM) cues underpin the perceptual experience of rhythm. This impairment is measurable in infancy for babies born in dyslexic families, who already show impaired discrimination of rise times ([Kalashnikova et al., 2018]).

[10] https://wearnotch.com

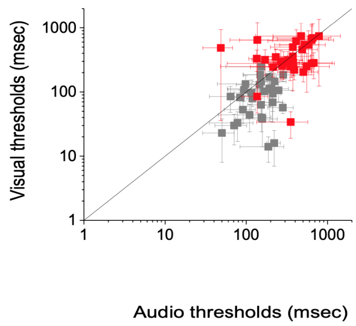
> Rise times also affect a person's experience of when in time a sound occurs – as rise times are the main cue to the 'perceptual centres' of sounds (old P-centres literature). My neural data for dyslexia suggest that the P-centre is perceived earlier in dyslexia – I mainly discuss this in terms of the speech signal, where we find atypical phase entrainment in the delta band (delta band = slower AMs) – e.g. [Power et al., 2013, Power et al., 2016]. Therefore, any auditory task with a temporal dimension is likely to be performed more poorly by children with dyslexia – such as your bisection tasks. In fact, your auditory bisection task is very similar to the format that we used in the musical metric task [Huss et al., 2011, Goswami et al., 2013], where we had one beat "out of time". This was a same-different judgement task, and the dyslexics could not hear which tune had the out-of-time beat. We have also shown rhythm problems in the motor domain – dyslexics are poorer at tapping in time with a beat. But neural work shows that this is because of the auditory modality being out of phase with the beat, not because of motor difficulties per se [Colling et al., 2017]. This also means that over developmental time, children with dyslexia are experiencing multi-modal information with one modality (audition) always "out of time" – this will impair the quality of information learned and have knock-on effects for multi-modal tasks. Specific to learning to read, children with dyslexia will read much less, therefore the brain will have significantly reduced experience with visuo-spatial processing e.g. following a line of print. This is particularly marked for languages like Italian with transparent letter-sound correspondences, where a serial processing strategy is very effective for reading print. The reduced visuo-spatial experience in processing linear arrays related to being dyslexia could in itself contribute to the impaired performance you find on the visual bisection task for Italian dyslexic children, as they are relatively old when you test them. Plus as noted, the visual bisection task has a temporal load. There are also lots of older studies showing a specific impairment in integrating auditory and visual information related to print in dyslexia e.g.[Snowling, 1980]. This is specifically an A-V impairment, not a V-A impairment."

In order to assess temporal performance in dyslexic children, we performed several tasks related to unisensory and/or multisensory temporal perception. Indeed, most of the past studies have investigated separately uni-sensory or multisensory sensory processing without considering them together. In particular, no studies to date have quantified the level of multisensory integration in dyslexic children considering the contribution of each uni-sensory signal on the multisensory processing. Our intention, led by IIT UVIP team, was to correlate the performance at such new psychophysical tests with the performance obtained at neuropsychological tests used to diagnose dyslexia and at classical temporal tasks previously presented in literature. To investigate the impact of sensory integration in dyslexic children, we measured in each child uni-sensory audio-visual temporal perception and their multisensory integration. A not-linguistic audio-visual temporal bisection task was performed in 32 typical and 32 dyslexic children. In

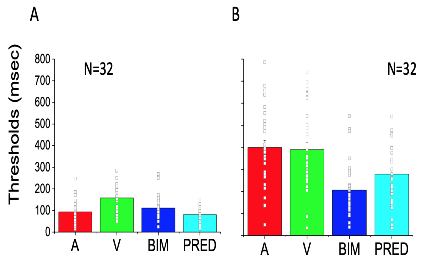
Figure 5.25: The bisection task



the task (Fig. 5.25), three different classes of stimuli (visual, auditory and audio-visual) were presented sequentially for a total duration of 1 second. Participants were required to indicate, by pressing a button, whether the middle stimulus appeared closer in time to the first or the third stimulus (temporal bisection). In the visual task (Fig. 5.25A) the participant was presented with a sequence of three lights: red, yellow and green. The subject had to determine whether the second (yellow) light appeared closer in time to the first (red) or to the last (green) light. Similarly, in the auditory task (Fig. 5.25B) the participant had to say if the second sound was presented closer in time to the first or to the third. In the bimodal task (Fig. 5.25C) the participant perceived a sequence of three lights associated with three sounds (like the uni-sensory stimuli). The visual and the auditory stimuli were presented either at the same time or "in conflict", with the auditory stimulus preceding or trailing the visual one. The procedure was identical to that of Gori et al. [Gori et al., 2012]: in the second stimulus the light preceded the tone by Δms ($\Delta = 0$, or ± 50 ms), while in the first and the third stimulus the offset was inverted in sign, so the light preceded the tone by - Δms. We varied the timing of the second stimulus (tone and flash together) to span the interval between the first and third stimuli. We used a child- friendly setup. The visual stimuli were 1° diameter LEDs displayed for 75 ms. Auditory stimuli were 750 Hz tones played for 75 ms. This temporal bisection stimulus was previously used to study temporal representation in typical children and in children with auditory deficits. Data were collected and analyzed by IIT UVIP team. Results were modeled by using the Bayesian approach and comparisons with reading skills and age of children were performed. Results suggested that children with dyslexia, compared with typical children, exhibited poorer unimodal thresholds requiring greater temporal distance between items for correct judgment. Multisensory thresholds were also higher than controls but interestingly they were well predicted by the Bayesian model. This result suggested that the multisensory deficit in dyslexia could be associated with an impaired unisensory audio-visual processing rather than on a multisensory integration deficit per se. Lower temporal skills also correlated with lower reading skills in dyslexic children suggesting a possible link between this temporal capability and reading abilities.



(a) **Individual thresholds for typical and dyslexic children obtained at the bisection task.**Individual visual thresholds against auditory thresholds for the typical group in gray and for the dyslexics group in red.

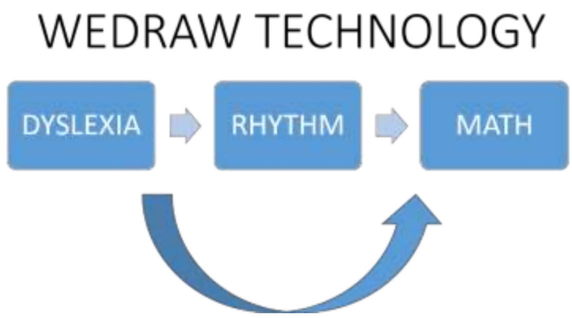


(b) **Average of unisensory visual and audio and bimodal thresholds obtained at the bisection task.** Average thresholds: visual, red; audio, green; audio-visual, blue; Bayesian prediction, light blue. A. Unisensory and multisensory average thresholds and individual data for the typical group. B. Unisensory and multisensory average thresholds and individual data for the group of dyslexics children. Error bars represent the standard error of the group.

Figure 5.26: Results in uni- and multisensory bisection task.

Several works indicate that dyslexia is linked to poor rhythmic abilities, mainly because phonological development is based on the accurate perception of metrical structures in language [Wood and Terrell, 1998], [Goswami et al., 2002]. For this reason, the partner from IIT UVIP assessed whether dyslexia is also associated with impairments in mathematical thinking, specifically with the ability to reason proportionally in terms of magnitude comparisons, probably because mathematical and rhythmic abilities share a common language. They also tested the children in a temporal reasoning tasks. Both the studies data are still in processing at the moment, but for the first study it seems to be a correlation between dyslexic children failing to perform the mathematical task and dyslexic children' performance in a writing words task in which they are asked to write known words (r = -0.6, p ¡ 0.05). Since proportional reasoning deals with the sequential comparison of rational quantities, similarly to what happens for rhythm perception when comparing consecutive temporal intervals, we speculate that mathematical thinking might be linked to rhythm perception. Indeed, it has been hypothesized that speech rhythm and music rhythm share common aspects defined by the metrical organization in language and music respectively [Huss et al., 2011], but no study to date has explored the link between dyslexia and

Figure 5.27: Main concept behind the development of new technology for dyslexia screening



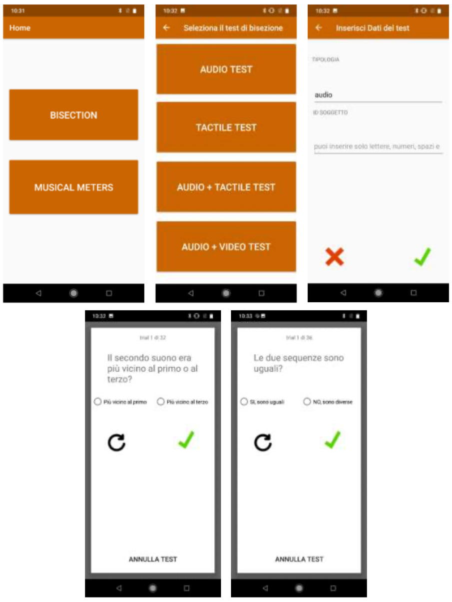
the metrical structure of mathematical thinking. weDRAW findings support the idea that dyslexia symptoms can be early identified by alternative non-reading tasks such as the proportional reasoning task used in this work. For what it concerned the second study, no difference seems to emerge between the performance of typical children and the performance of dyslexic children.

5.4.1 Dyslexia screening app

Having said that, in the framework of the work done to improve dyslexia screening test, we developed, in partnership with IIT UVIP, a new tool app containing two tests for screening dyslexia. The app was developed within Android libraries and built on smartphones, to be able to easily manage the test activity. The idea on which the app design was based was to be able to develop a tool useful in several everyday contexts, to be used easily by researchers but also teachers and parents, with a high usability profile (see Fig.5.27). The two tests included in the app are: the bisection test as presented above in this section and the musical meters task, as suggested by the expert Prof. Usha Goswami and explained in a recent article [Huss et al., 2011].

The first version of the app is freely distributed in the PlayStore, including only the bisection test. The second version of the app has been developed by the in order to include both the bisection and the musical meter tests (see Fig. 5.28). If participants select the button "bisection", the bisection test will start and the first screen of the app allows the users to choose the sensory modalities with which they want to do the test (audio, tactile, audio+tactile, audio+video). Then participants are required to insert their personal information (id_soggetto) and after final confirmation the test will start. The test includes 20 trials, during which both congruent and incongruent stimuli are presented. On each trial, the app asked the user to identify if the second stimulus provided by the system was closer to the first or the third. A txt file with the results of the test is produced after every use of the app. If participants select the button "musical meters", the musical meters test will start and the first screen of the app allows the user to insert their personal information

Figure 5.28: Screenshots from the new app to screen for dyslexia



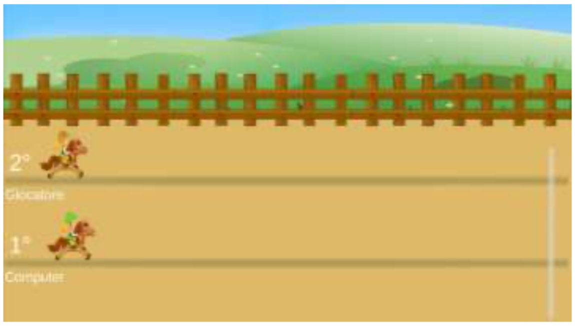
(id_soggetto) and after final confirmation the test will start. The test includes 36 trials, during which both same and different trials are presented. On each trial, the app asked the user to indicate whether the two sound sequences presented consequentially are identical or different. A txt file with the results of the test is produced after every use of the app.

5.4.2 Dyslexia training app

A second application developed in collaboration with IIT UVIP consisted in a mobile game on rhythmic tapping. The game, called “The synchrohorse race” (Fig. 5.29), was created again for mobile systems, to be able to easily reach users. The game is a 2D Unity game based on the idea of a horse race, in which to move the own horse the only thing to do is to correctly tap on the screen the tempo of some music excerpts of less of 20 seconds. The music pieces were selected to be representative of several musical tempo, with different BPM, and the game randomly selected the list of pieces to reproduce, in a way that it’s really unlucky that the same pieces can be listened more than once, if the game is played a couple of time in a row. The game is intended to be a first trial of adopting the weDRAW research framework and developing new tools to help children with dyslexia to enhance their temporal skills.

At the moment, we tested the app “The synchrohorse race” in 10 dyslexic children, in order to gain insights into the role played by rhythm perception in shaping reading performance. Children have been enrolled in order to perform a 10-days training during which they will play the

Figure 5.29: The synchrohorse race app to train rhythmical abilities in dyslexic children, based on the idea that rhythm perception could be linked to reading processing.

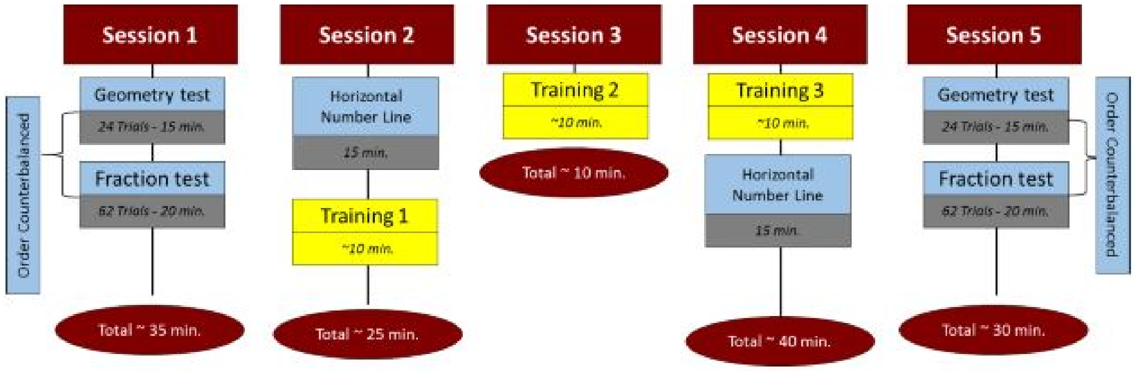


game for 20 minutes a day. Before and after the training they perform two tasks meant to test their temporal perception. In the bisection task, as reported above, they are required to indicate whether the second stimulus presented is closer in time to the first or to the third one presented. In the musical meter task [Huss et al., 2011], they are required to indicate whether two consecutive melodies are the same or different depending on changes in the meter perceived. The data of this data collection are still in a processing phase, carrying on by our partner from IIT UVIP team, and we hope to be soon able to understand if this training was successful.

5.5 Evaluation

Since novel methodology of the weDRAW approach was grounded on the idea that our sensory systems can play different roles in learning specific concepts during development, all these different game activities provided auditory feedback associated to the visual input. Moreover, all of these serious games and activities used an embodied and enactive approach for learning to store knowledge in the form of motor responses acquired by the act of doing. Our partner from IIT UVIP examined mathematical knowledge across different sensory modalities in children across different ages by using new hardware and software platforms to support testing with the serious games. They defined an evaluation of mobility and rhythmic skills in children to evaluate technology and serious-games on arithmetic and geometrical skills both in sighted and visual impaired children. Moreover, since we think that usability evaluation is itself a core concept to deepen in multimodal educational technology design, starting from few findings from the literature, we developed a new design usability test to evaluate technology suitability for both typically developed and visually impaired children. In next sections, we will deepen weDRAW activities evaluation from both these point of view.

Figure 5.30: Methodological approach for the psychophysical validation of the AngleShapes-Game and MusicFractionsGame activities.



5.5.1 Psychophysics evaluation

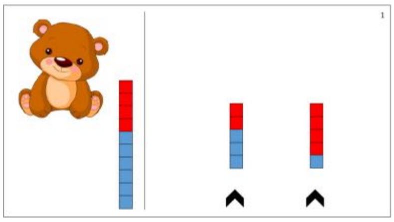
Tests are defined to measure improvement, at the cognitive and perceptual levels of those concepts, following training with the serious games in children. The goal of the study was to validate the AngleShapes Game and the MusicFraction Game activities that was designed to convey mathematical and geometrical concepts in a non-formal manner during primary school years. The core aim of the games were to provide an educationally robust but engaging vehicle to teach core complex concepts such as angles, shape transformation and fractions.

Children between 6 and 9 y.o. (n = 24) participated in the game validation. They were divided into two groups: training with AngleShapes Game (n = 12) and training with MusicFraction Game (n = 12). Both groups underwent five activity sessions: in the first session they were administered with the first part of the pre-test; in the second session, they were presented with the second part of the pre-test, introduced with the task and then performed it; in the third session they performed the task; in the fourth session, they performed the task and then were tested with the first part of the post-tests; in the fifth session, children were tested with the second part of the post-tests. This organization is represented in Fig. 5.30. To evaluate the efficacy of the AngleShapes Gmae and MusicFraction Game activities, we tested children's knowledge about concepts such as proportional reasoning, numerosity and general geometrical knowledge.

5.5.1.1 Measure of Proportional reasoning

Proportional reasoning involves understanding the multiplicative relationships between rational quantities (a/b = c/d), and is a form of reasoning that characterizes important structural relationships in mathematics and science, as well as in everyday life. In the Proportional Reasoning Task, three participants were asked to select a proportion that matches a target juice mixture,

Figure 5.31: **Proportional Reasoning Task (trial example)**. On each trial, participants are required to indicate which of two alternatives (right panel) best represents the target juice (left panel) representing a mixture of water/blue and juice/colored parts.



where proportionality was determined by the relative quantities of juice and water parts (see Fig. 5.31 for an illustration).

5.5.1.2 Measure of numerosity

Several studies indicate that numerosity competence is organized around a mental number line, in which number symbols are represented non-verbally in an ordered spatial organization. IIT UVIP developed a revised version of the classical number line task, [Siegler et al., 2011], [Barth and Paladino, 2011], [Aagten-Murphy et al., 2015], that assesses knowledge of positive and negative numbers in the horizontal and the vertical dimensions. On each trial, participants were required to localize on a bold horizontal number line from 0 to 100 (or from -100 to 0) the position of a specific number (e.g. 17) and localization errors were measured as the distance in numbers between the target and the indicated position.

5.5.1.3 Measure of general visuo-spatial abilities related to geometry

For this evaluation, IIT UVIP referred to a validated Italian battery of tests specifically developed to link visuo-spatial abilities with geometrical knowledge [Mammarella et al., 2012]. Since our interest here was to test potential perceptual improvements given by either one of the two training proposed to children, we focused on the battery subset of tests that evaluate visuo-spatial abilities in primary school children.

5.5.1.4 Training

AngleShapesGame activity. As presented before, 5.2.4.1, in these activities children were required to build a house by making angles with their body, i.e. by moving their arms. Children were asked to reproduce with their arms each angle that composes the target figure. In the first training session, the target figure is a square composed of four 90° angles. In the second training session, the target figure was a triangle, composed 2 angles of 45° and one angle of 90°. In the third training session, children were asked to reproduce 11 angles to compose the house in the following order: firstly the façade of the house (composed by 4 angles of 90°); secondly, the roof of the house (composed by 2 angles of 45° and one angle of 90°); thirdly, the door of the house (composed by 4 angles of 90°). The target figure was always represented on the screen in front of the participants. Each angle to be composed was signalled by the interface. Once the angle was made by the child, s/he was asked to confirm the angle by maintaining the arm orientation for 10 seconds. Were the response correct, the aperture between the arms would turn green and the response taken to continue with the game. If the angle was not correct, the aperture between the arms would become red and the response not taken until the child reproduces the correct angle. In order to facilitate the comprehension of angles, a sound cuing the aperture of the angle was added to the game.

The MusicFractionsGame activity. The fraction group (Fig. 5.32b) performed an activity that used rhythm and body movement to learn fractions. Participant's body movement was recorded by means of Kinect.

On each training session, the child was standing up in front of the Kinect such that his body movement could be tracked and s/he was presented with a screen to show game visual information and sounds emitted by two speakers. The training was divided in two phases, guided exploration and target phase. The guided exploration was presented only at the first training session where the child was asked to familiarize with the game rationale: fraction's denominator was defined by legs widening while the numerator was defined by the distance between hands.

(a) Representation of a child performing the AngleShapesGame activity training

(b) Representation of a child performing the MusicFractionsGame activity training

The child was presented with a circle that was turned in a pie chart as s/he widens the legs. The number of slices composing the pie chart represented the denominator. This information was also transmitted through an auditory feedback for which a strong sound was followed by a number of weaker sounds to indicate the denominator (e.g. denominator = 4, corresponds to 1 strong sound and 3 weak sounds). In details, the experimenter told the following instructions to the child:

Try to widen your legs to divide the cake. Try dividing the cake into two slices.

> Note that there are two different sounds: one louder and one weaker, one after the other. Now try dividing the cake into another number of slices, for example, make 4 slices. What do you feel? One strong sound and three weak sounds. Pay attention to the sound in this game, because you need to count how many slices you divide the cake into. Now divide the cake into two slices again. At this point, the child is asked to confirm the selected denominator. In details, when the child opens his legs sufficiently, i.e. to divide the cake into 2 slices, s/he is told to stand still and confirm with his open hand and outstretched arm, facing the Kinect. He must stand still for a few seconds. Thus, the experimenter tells the child the following: "To confirm the number of slices, hold still for a few seconds in this position - do not move your legs and use your outstretched hand to press the confirmation button you see on the screen.

Once the denominator has been defined, the child was asked to make a half (i.e.,1 / 2), that was to take half of the cake as highlighted by the visual feedback (i.e. the selected slice(s) change colour) and the auditory feedback (a number of bell rings equal to the numerator). To this aim, the experimenter provided the child with the following instructions:

> at this point move your arms to take a slice of the cake. When you have selected it, and the slice will light up, stay still for a few seconds to confirm. You will hear a new sound, a bell, that will tell you how many slices of cake you have taken.

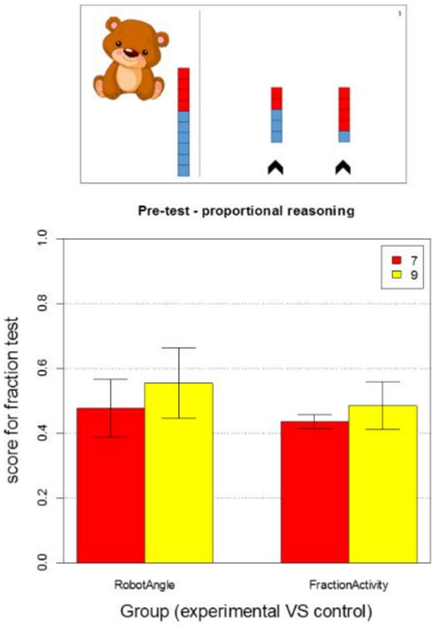
At this point the child was asked if s/he had understood how to interact with the game and s/he was asked to perform another fraction (i.e. 1/6). When the child selected the correct number of slices, the experimenter checked that the child recognized the sound associated with the selection, i.e. they heard one strong sound and five weak ones. Once the child had understood how to interact with the game, the experimenter moved on to the next step. In the target phase, the experimenter defined the fraction to be done by the child, first selecting the denominator, which could reach a maximum of 6, and then the numerator which could be adjusted accordingly, also up to a maximum of 6. The target fractions presented to the children are 3 per session, in details:

- First session: 1/4, 3/3, 3/5;
- Second session: 1/2, 3/4, 3/6;
- Third session: 2/5, 5/6, 2/4;

5.5.1.5 Proportional reasoning results

The results for the proportional reasoning were represented by focusing on subjects' performance calculated as the proportion of correct responses in the test for each subject and averaged for each

Figure 5.33: Results of the pre-training



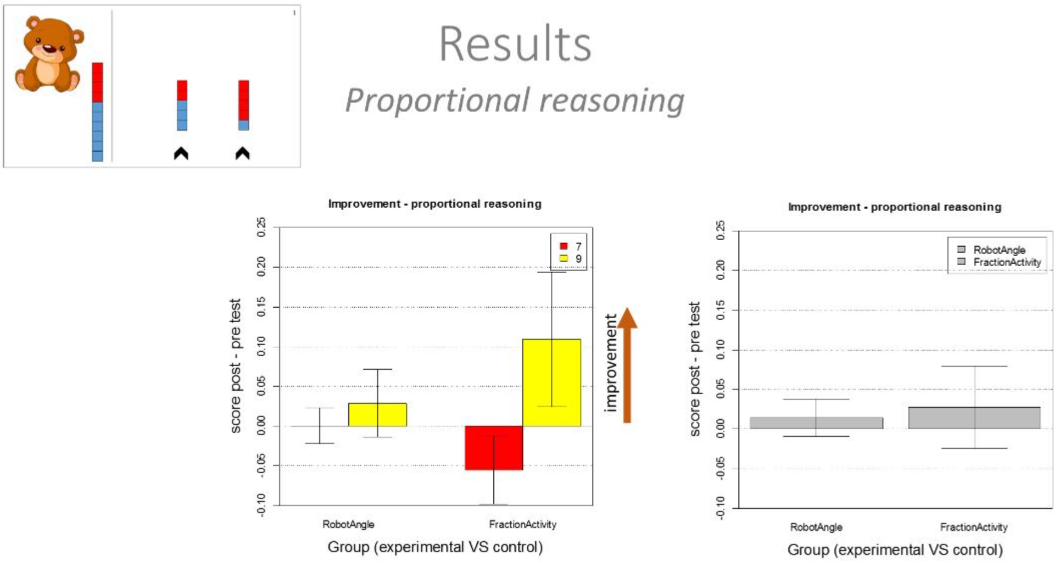
group. In the pre-test phase, we could observe a similar pattern of results between the two groups (experimental and control), and across age groups (see Fig. 5.33).

Comparison of performance to the post and pre-test measurements was indicated as delta (calculated as post-pre test), where a positive delta indicated an improvement, whereas a negative delta indicated a worsened performance. As represented in Fig. 5.34, the group of 9 y.o. after the experienced training with the Fraction activity. Children of 7 y.o. instead, did not show such improvements.

5.5.1.6 Measure of numerosity

The results for both the negative and the positive horizontal number line were represented by focusing on the error in indicating the point on the line corresponding to the target number. Mean of the absolute error across trials was calculated for each subject and averaged for each group. In the pre-test phase, we could observe a similar pattern of results between the two groups (experimental and control), thus indicating a decreasing error with age, that was older children were better in the task compared to the younger peers (see Fig. 5.35). Comparison of performance to the post and pre-test measurements was indicated as delta (calculated as post-pre test), where a negative delta indicated an improvement, whereas a positive delta indicated a worsened performance. As represented in Fig.5.36-5.37 both groups of 7 y.o. participants who experienced the training with the AngleShapesGame or with the MusicFractionsGame activities,

Figure 5.34: Results of the effect of the training



showed a better performance after the experienced training for the negative and the positive number line test. Children of 9 y.o. instead, did not show such improvements.

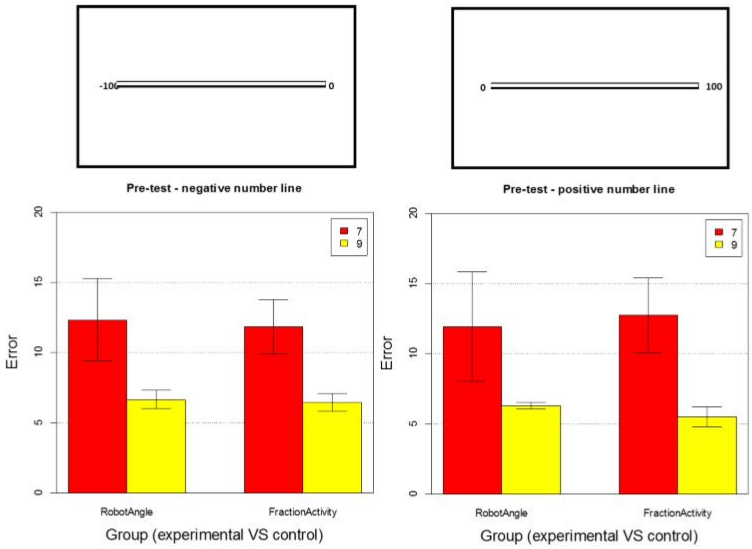
5.5.1.7 Measure of general visuo-spatial abilities related to geometry

The results for the visuo-spatial abilities in primary school children were represented by focusing on the proportion of correct responses across trials (6 exercises, 24 trials in total). Mean of the proportion of correct responses across trials was calculated for each subject and averaged for each group. In the pre-test phase (Fig. 5.38), we could observe a similar pattern of results between the two groups of children (7 years old and 9 years old), thus indicating a constant performance across development.

Comparison of performance between the pre- and post measurements was indicated as delta (calculated as post-pre test), where a positive delta indicated an improvement, whereas a negative delta indicated no improvement. As represented in Fig. 5.39 and 5.40, only 9 years old participants who experienced the training with the AngleShapesGame showed a better performance after the training, while 7 years old participants did not show such improvements.

Results indicated that children of 7 y.o. children improved their performance in the number line task with both training activities (i.e. AngleShapesGame and MusicFractionsGame activities) and 9 y.o. children improved their performance in the geometrical task after the training with AngleShapesGame. Such age specificity might depend on the fact that the youngest group of children was less familiar with the concept of numbers along a continuum. Both trainings indeed

Figure 5.35: Results of the pre-training



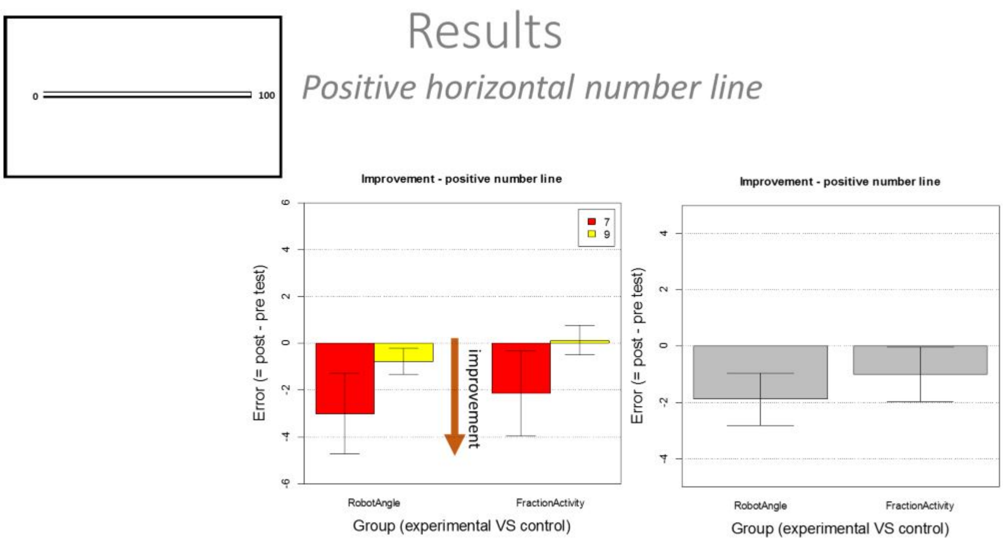


Figure 5.36: Results of the effective training on numerosity on Positive Number Line

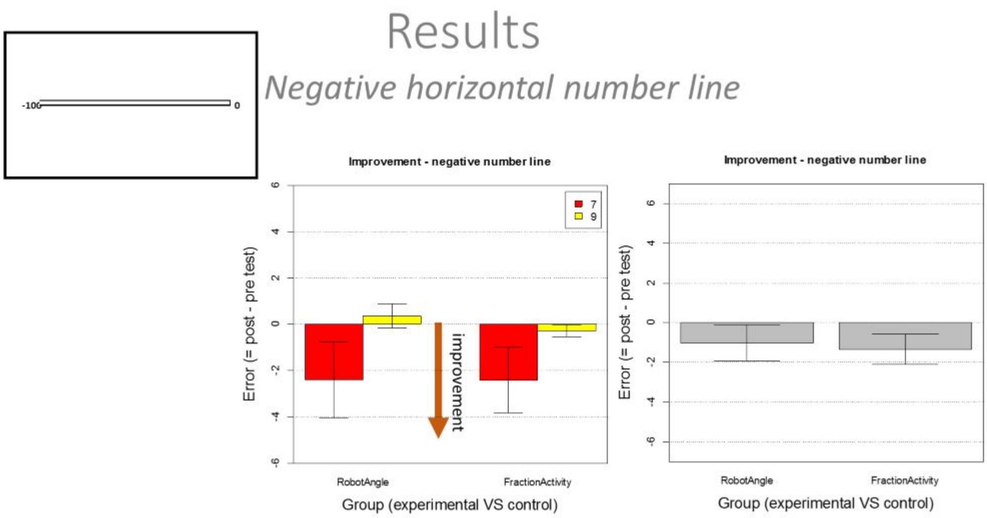


Figure 5.37: Results of the effective training on numerosity on Negative Number Line

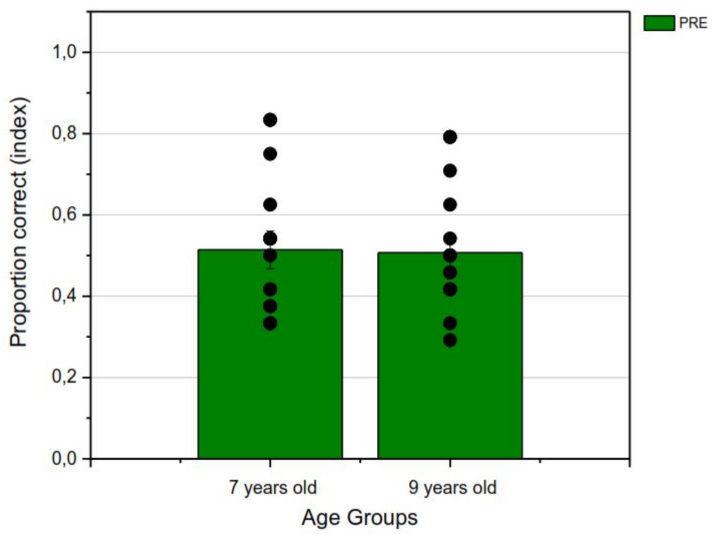


Figure 5.38: Pretest assessment of visuo-spatial skills in 7-to-9 years old children

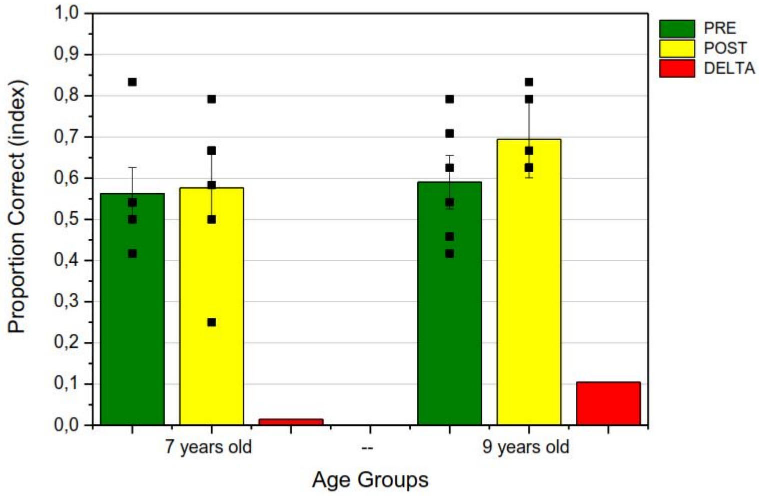


Figure 5.39: **Performance before and after the training with the AngleShapes-Game.**Compared to 7-years old children, 9-years old children show to perform slightly better after performing the AngleShapesGame.

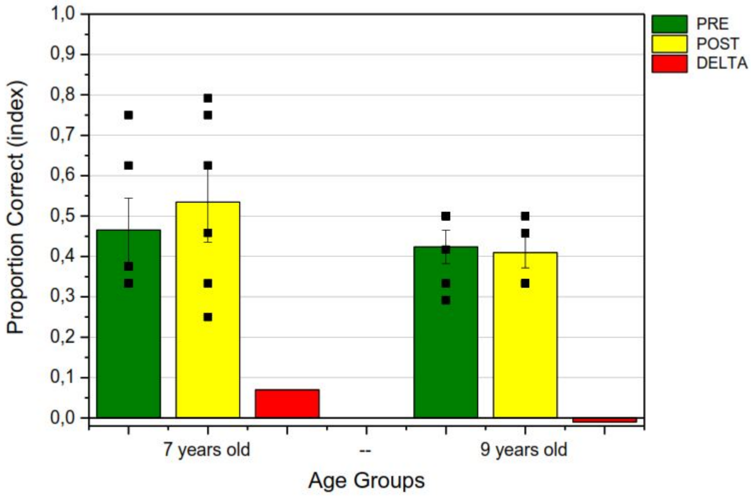


Figure 5.40: **Performance before and after the training with the MusicFractions-Game.**While 9-years old children don't show any specific improvement after the training, 7 years old children show to perform slightly better after performing the MusicFractionsGame activity.

made them more familiar with the relationship between numbers and increased quantities represented as angle apertures corresponding to sound and arm apertures (i.e. the AngleShapesGame activity) as well as increasing and decreasing numbers corresponding to arm/legs apertures and rhythm (i.e. in the MusicFractionGame activity). Interestingly, in the proportional reasoning test only children of 9 y.o. who experienced the fraction activity training showed improvements. This result might be related to the school level and learning stage of the selected groups of children. The concept of fraction was indeed taught during or after the 3 level of primary school. Moreover, the fact that 9 years old participants who experienced the training with the AngleShapesGame showed a better performance after the training in the geometrical and symmetry tasks, while 7 years old participants did not show such improvements, indicated that depending on age the effects of the training could be different. Therefore, children of 9 y.o. might benefit the most from the proposed training in the context of proportional reasoning and fraction understanding as they already had the basis to learn such concepts.

5.5.2 User multimodal technology evaluation

The widespread of gaming technology has impacted daily routine of children and adolescents, going beyond entertainment purposes [Mayo, 2007], [Cheng et al., 2015], and the projects on which this dissertation is based, are a clear example of this tendency. In this context, serious video games play a valuable role in the development and enhancement of diverse types of cognitive skills [Rocha et al., 2016], as well as for teaching and learning purposes [Cheng et al., 2015], including people with multiple types of disabilities [Cheng et al., 2015], [Sánchez and Olivares, 2011], [Durkin et al., 2015]. Furthermore, visually impaired students have been using serious multimodal video games based on audio and haptics to foster mental skills, such as logical reasoning, navigation, mental mapping, and spatial cognition [Connors et al., 2014], [Lahav and Mioduser, 2008b]. Such video games are designed to also help blind people to transfer virtually acquired skills to everyday life [Connors et al., 2014]. However, the development and enhancement of the intended cognitive skills will be possible through these games only if they manage to combine the adequate modalities, while carefully coordinating interface and feedback to represent abstract information [Sanghvi et al., 2011], [Cheng et al., 2015], [Lahav and Mioduser, 2001]. As we saw in previous sections of the dissertation, for visually impaired students, children in particular, game modalities must afford a precise interpretation of the information conveyed [Bernsen and Dybkjær, 2009]. Consequently, usability is fundamental in this context, especially considering that digital games usually require constant interaction, and focusing on usability issues rather than on learning would be frustrating and undesirable [Bellotti et al., 2010], [Ardito et al., 2006]. Nevertheless, usability evaluation of multimodal serious games for visually impaired learners lacks reasoning, regarding what game aspects to evaluate and how to proceed the assessment [Sánchez et al., 2015]. Evaluators conducting usability tests involving visually impaired people should keep in mind that traditional Usability Evaluation Methods (UEM) [Hartson et al., 2001] are designed for users without disabilities [Chandrashekar et al., 2006].

Moreover, multimodality adds further complexity to this scenario since specific issues differentiate multimodal usability evaluation from the evaluation of traditional user interfaces, such as GUIs [Bernsen and Dybkjær, 2009]. Besides, usability evaluation of multimodal games involving young visually impaired and blind learners has also to take into account the differences between children and adults with the same condition. Blind and low-vision children, in fact, cannot fully perceive anything at once; instead, everything has to be constructed [Spencer et al., 1992]. In addition to that, they are still learning and experiencing situations, constructing cognitive schema, that they can differ from those of adults, according to their linguistic, intellectual, and motor skills reached level [Raisamo et al., 2006]. To the best of our knowledge, few works addressed usability evaluation of multimodal games for blind and low-vision children. In a systematic literature review, Sánchez and colleagues [Sánchez et al., 2015] analysed 25 papers describing evaluation and design of multimodal video games and virtual environments for cognitive enhancement of blind people. The authors discussed in details how studies with similar goals followed different procedures to conduct usability evaluation, identifying a lack of research in this regard. They remarked that some studies made unconfirmed assumptions about ease of use, learnability, and interaction of these games, as they do not perform usability evaluation involving potential users [Torrente et al., 2009], [Trewin et al., 2008], [Lahav and Mioduser, 2008a]. Darin and colleagues [Darin et al., 2015] proposed a *Standard List of Usability Problems* (SLUP) based on the analysis of the usability evaluation reports of five target multimodal video games for cognitive improvement of visually impaired children [Darin et al., 2017]. SLUP contains 61 issues related to the interface and interaction features that commonly impact the multimodal gaming interaction of blind learners and aims to help designers avoid recurrent usability issues regarding audio, adaptation, interaction mode, and feedback in the design of such games. Comparing data gathered from videotaped user observations applying Thinking Aloud Protocol (TAP), videotaped interviews, and answers to a questionnaire, the authors identified that SLUP could be used as a ground to develop specific usability evaluation instruments. To develop a technology evaluation meaningful in the framework of weDRAW, we considered as a started point of our research both SLUP and *Checklist for Usability Evaluation of Multimodal Games for Children who are Blind* - CLUE, a technology evaluation checklist developed by Darin and colleagues in 2018, [Darin et al., 2018], starting from the findings of SLUP. Multimodal serious-games aimed at developing and enhancing cognition in young visually impaired learners, can be described accordingly to four dimensions: interaction, interface, cognition and evaluation [Darin et al., 2017], [Darin et al., 2018]. These aspects indicate core elements of game interaction and interface characterization, along with the cognitive process meant to be developed and enhanced, and the type of evaluation implemented. According to this characterization, the interface and interaction features that most impact multimodal digital games for visually impaired learners are Audio, Adaptation, Interaction Mode and Feedback [Darin et al., 2017]. We enriched the description of such features by aggregating physical carriers, according to Bernsen's [Bernsen and Dybkjær, 2009] modalities taxonomy. In Table 5.41 are presented the modalities usually present in multimodal serious-games for visually impaired people.

MEDIUM	MODALITY CHARACTERIZATION	
GRAPHICS	Interface	Bidimensional images, maps or graphs Tridimensional images, maps or graphs Written text
	Feedback	Contextual visual cues (using graphic interface features)
	Adaptation	Size Contrast Color Scheme
ACOUSTICS	Interface	Spoken audio Speech synthesis Iconic sounds Spatialized sounds Stereo sounds Abstract earcons
	Feedback	Contextual audio cues (using sonorous interface features)
	Adaptation	Speed Intensity
HAPTICS	Interface	Tactile Kinesthetic
	Feedback	Force Vibration Pressure Motion
	Adaptation	Intensity Frequency

Figure 5.41: Usual modalities for multimodal digital games for visually impaired learners [Darin et al., 2015].

The CLUE checklist was designed to be used by researcher, practitioners, and teachers during field tests involving visually impaired children interacting with video games. It consists of a 40-item checklist, addressing multiple aspects of gameplay and multimodality, including acoustics, graphics, and haptics. CLUE aims to help in the identification of the real problems that affect the multimodal gaming interaction of these users, in a practical way. Starting from that checklist, we developed our checklist model here presented in the Appendix B.1. The checklist was developed to be used during the last experiments in schools and rehabilitation center that we did to evaluate the final weDRAW serious-games and activities with primary school children. It is intended to by used by researchers themselves or external observer that may be presented as well as CLUE. Furthermore, considering the weDRAW technologies is divided in three sections: (a) Gameplay; (b) Acoustics; (c) Graphics. It is composed by 38 items, that aimed at checking if the player had difficulties to accomplish specific tasks or understand feedback, or if she or he demonstrated to feel bored or stressed or confused and so on. The checklist for each item required to indicate with "Yes" or "No" if the specific item happened, in which particular tasks and with a 4-items Likert scale with which severity. The questionnaire was intended to be used both in primary schools with the whole classrooms, and in individual tests carried out at Istituto Chiossone, the rehabilitation centre for visually impaired people based in Genova and partner of the weDRAW consortium. The idea to apply the usability evaluation to both groups of users is due to the focus of the weDRAW project to develop unique technology solutions suitable to be used by both typically developed and visually impaired children. In the primary schools of Genova, 111 children were tested with the AngleShapes Game. Mean age of the children was 8,66, with a gender distribution of 2,13 for female students and 1,88 for male students. The evaluation was carried out by one external psychologist and one researchers. As showed in Fig. 5.42, typically developed children had issues especially with acoustics component of the game. Speaking of that, we had to say also that observers annotated for more than 90 questionnaires that the usability issues they signed for the acoustics component were due, for the majority, to the lack of attention to the sounds of the game. This can be understandable, considering the priority that for typical children vision has. Unfourtunately we were unable to test the game with a statistically significant population of visually impaired children. We had just the change to test it with 3 low-vision children, all female students, mean age 10 years old. In Fig. 5.43, usability issues resulted from the questionnaire are reported. As it can be seen, for visually impaired children gameplay and acoustics did not represent an issue, while graphics component did, with high severity value for at least one child. Of course, these results are not statistically significant, but they represent just an indication of further work needed in clarify the game interaction and the relation between feedback modalities to be suitable for as many users as possible. It is interesting to consider how much the same information conveyed by the same sensory modality, such as the acoustic one, has been correctly used or misused or just not considered by the children, basing of the presence of the visual channel information.

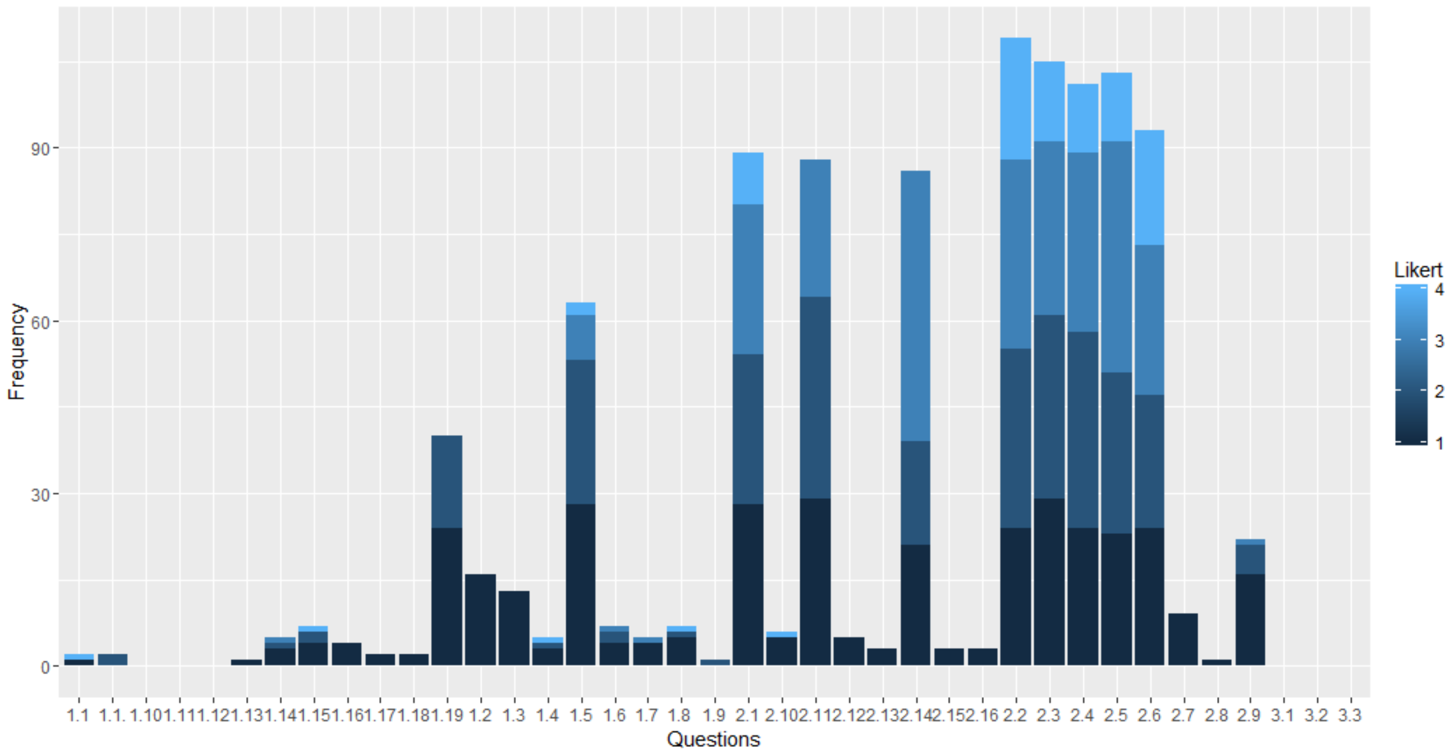


Figure 5.42: Frequency annotation of usability issues on AngleShapes Game for typically developed children

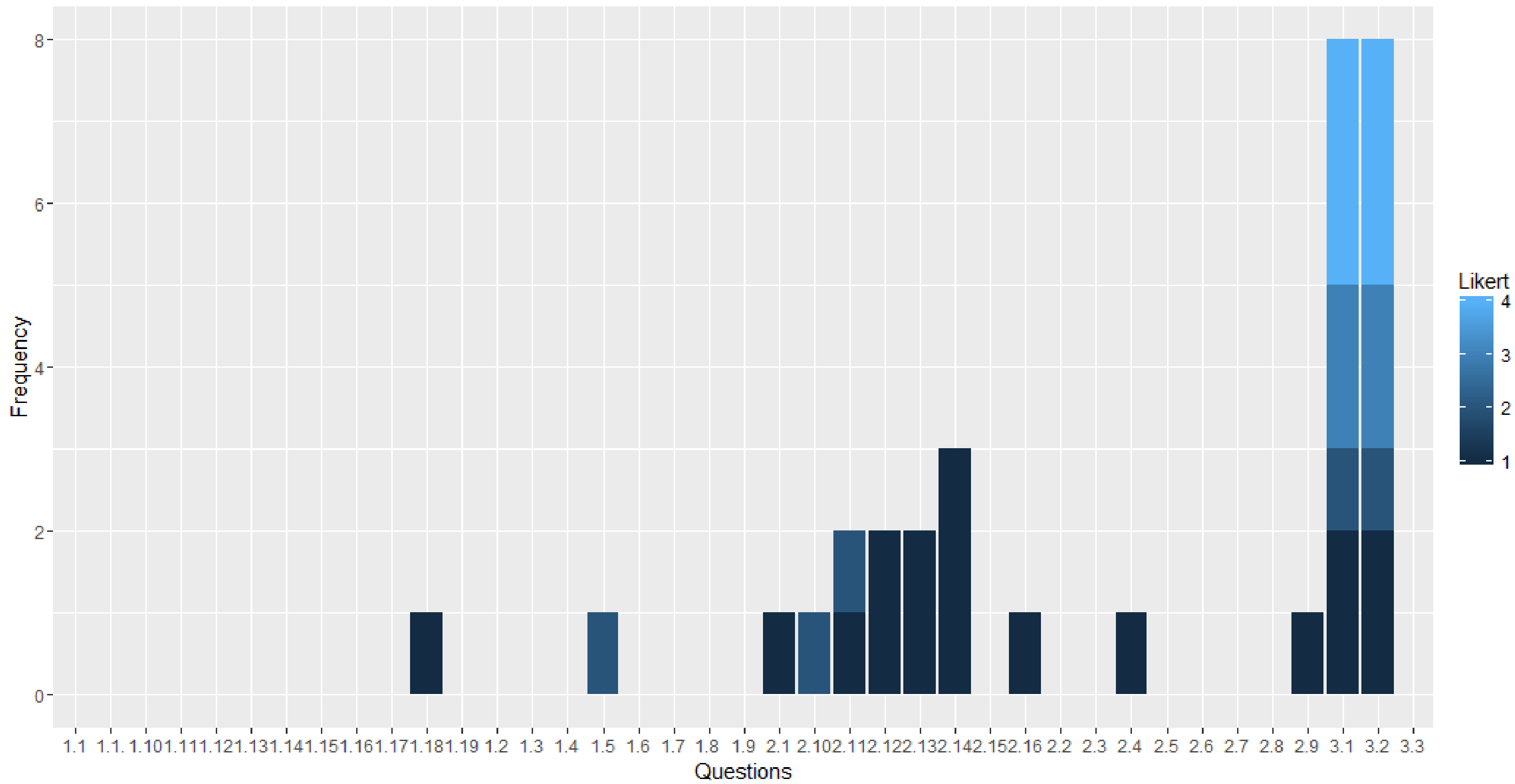


Figure 5.43: Frequency annotation of usability issues on AngleShapes Game for visually impaired children

Chapter 6

The TELMI Project

6.1 Overview and objectives of the project

Playing a musical instrument is a highly complex activity. It requires a complex combination of mental and sensorimotor skills, which are acquired during a long learning trajectory. Music learning is mostly based on a master-apprentice model in which the student observes and imitates the teacher, the teacher provides verbal feedback on the performance of the student, and the student engages in long periods of self-study without teacher supervision. However, learning under the master-apprentice model is difficult because the time lag between the student‘s performance and the teacher‘s feedback makes this last to be dissociated from the online proprioceptive and auditory sensations accompanying the performance – this is especially relevant since most of student‘s performance practice takes place long after the teacher‘s feedback. The resulting long periods of private-study the students spend frequently make the learning of musical instruments a rather harsh and solitary experience, resulting in high abandonment rates. In addition, approaches to teaching the biomechanics of musical performance may be based on subjective and vague perception, rather than on accurate understanding of the principles of human movement. The student requires to mirror the teacher‘s body language, according to verbal feedback which is often susceptible to ambiguous interpretation. Other factors such as the belief in the need for long hours of repetitive practice to acquire technical skills or the competitive nature of music performance often result in inefficient practice with little consideration for the efficacy. For the student, consequences of such an uninformed pedagogy can range from a frustrating lack of progress, and consequent abandonment, to health problems. Musical performance shares many characteristics in common with other skill-oriented activities. For example, commonalities between sports and musical performance are obvious, particularly in the area of biomechanics. It is reasonable to postulate, therefore, that some methodologies already used successfully in sports kinesiology could be useful in studying aspects of musical performance. These include motor learning theory, learning models, and the use of technology to analyse and validate the effects of training.

The TELMI -"Technology Enhanced Learning of Musical Instrument Performance" addressed these issues aimed at designing and implementing a new multimodal interaction paradigm for music learning, to develop assistive, augmented feedback complementary to traditional teaching.

6.2 Review of pedagogical methodology in violin learning

Self-regulated learning is a model that outlines the importance of effective planning, execution, and review strategies to foster efficient learning, including the development of meta-strategies that require knowledge of the nature and benefit of each component and an ability to use each when appropriate. Examples of such strategies include goal setting (planning), mental rehearsal (execution), and error detection (review). Violinists learn their instruments in specific contexts and have a vast set of resources at their disposal. In these situations, however, musicians also face major challenges, including barriers to two-way communication between student and teacher, the competitive nature of advanced musical study, and a restrictive environment based on traditional education models that can be resistant to change. One of the objective of TELMI Project was exactly to drive this change of prospective from traditional violin pedagogy to a technology-based one. Violinists must learn a highly complex set of biomechanical techniques and musical concepts that need to be integrated and coordinated for effective performance. These include skills related to body position, handling of the instrument, and techniques of the right (bowing) and left (fingering) hands. Hundreds of years of often conflicting pedagogical literature outlines the physiological, psychological, and artistic methods and goals in approaching each technique. Some technological tools, such as online video resources, videoconferencing capabilities, and standard audio-video recording devices, are growing in availability. Numerous advanced feedback systems tailored to the violin have been developed but have not seen integration in the music community.

6.2.1 Self-regulated learning

Self-regulated learning refers to processes whereby individuals assume personal responsibility for and control of their acquisition of knowledge and skills [Zimmerman, 1990]. It occurs when learners become "metacognitively, motivationally, and behaviourally active participants in their own learning process" [Zimmerman, 1989], which is in contrast to those situations where learners follow or merely react to external impetus or instruction. Three processes are central to effective self-regulation:

- forethought: the thought processes and personal beliefs that precede efforts to engage in a task;

- performance/volition control: processes that occur during learning that affect concentration and performance;
- self-reflection: the learner's reaction and subsequent response to the experience [Zimmerman, 1989].

Paris and Winograd [Paris et al., 1990] proposed that the regular self-assessment of learning processes and outcomes promotes more effective monitoring of progress, facilitates the identification and correction of mistakes, and enhances feelings of self-efficacy, which is the belief in one's ability to perform domain-specific skills [McCORMICK and McPHERSON, 2003], [Bandura, 1997], [McPherson and McCormick, 2006], [Pajares, 1996], [Ritchie and Williamon, 2012]. Although self-regulation has been studied predominantly within educational psychology, music-specific research has shown that musicians also benefit from taking an active role in initiating learning, in choosing an appropriate and effective pathway to develop their skills, and in systematically managing learning processes (see [Ritchie and Williamon, 2012]). To assist musicians in applying principles of self-regulation to their own learning, Jørgensen [Jørgensen, 2004] proposed four broad types of practice strategies, three of which correspond to the three central processes of self-regulation, plus a fourth for organizing one's practice strategies:

- planning and preparation strategies (i.e. 'forethought' in the self-regulated learning framework): for activity selection and organization, setting goals and objectives, and time management;
- executive strategies (i.e. 'performance and volition control'): for rehearsal, distribution of practice over time, and preparing for a public performance;
- evaluation strategies (i.e. 'self-reflection'): for process and product evaluation;
- meta-strategies: knowledge of strategies, and control and regulation of strategies.

By explicitly highlighting meta-strategies, Jørgensen encourages musicians to attain a thorough understanding of their own repertoire of practice strategies, together with the ability to control, regulate, and exploit that repertoire (see Fig.6.1).

Following Jørgensen's [Jørgensen, 2004] model of music-specific processes, in this section we summarize individual components of planning and preparation, executive strategies, and review and evaluation.

6.2.1.1 Planning and execution

Planning how practice time will be used is critical to the self-regulation process as it ensures the efficient and effective use of time and that short- and long-term goals are met. While the technol-

Figure 6.1: Model of self-regulated learning as a cycle of planning, execution, and evaluation practice strategies guided by overarching meta-strategies.

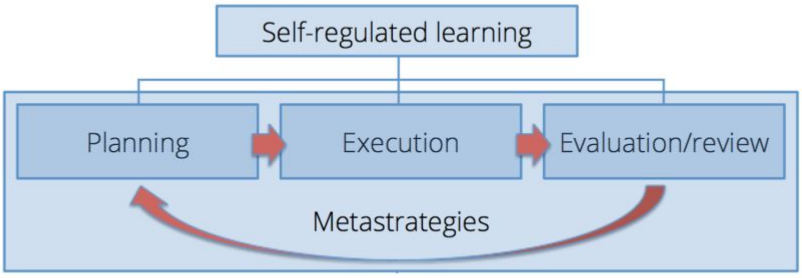
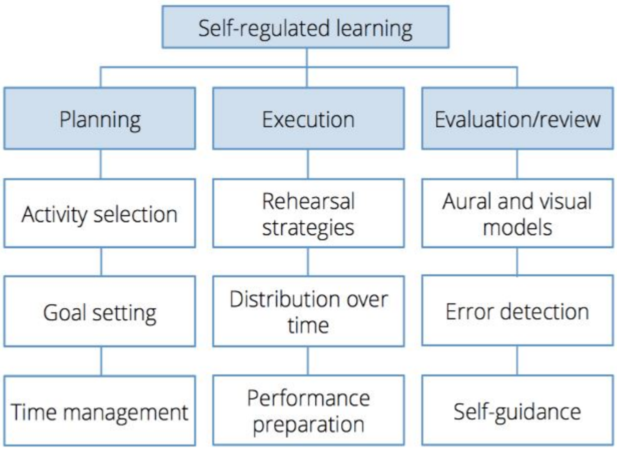


Figure 6.2: Components of the three self-regulation processes.



ogy that TELMI designed were aimed at given to musicians unprecedented amounts of feedback to evaluate their practice, the principles of self-regulated learning also demand that this information be used to decide how future time in the practice room will be spent. Thus, the technology developed should take an active role in guiding students and teachers through this process. This includes choosing (or recommending) what activities will be accomplished, what the goals are, and how time will be managed. How this planning is executed in practice relied heavily on what tools implemented and how the feedback systems inherent to the TELMI system may be used to guide real-time rehearsal strategies. Choosing what activities will make up a practice session is a key component of effective planning. The first level is to remove activities not related to learning (e.g. spending time on social media) and minimizing time spent in necessary tasks not related to learning (e.g. repairing an instrument). Within learning time, musicians may spend time in "playing practice" or "non-playing practice", which can include score study, making notes, reviewing recordings, and so on [Jørgensen, 2004]. This both maximizes opportunities for review and reflection within the session while also providing space for mental and physical rest. Musicians normally follow a pattern of warm-up exercises, technical exercises and études, and repertoire [Jorgensen, 1998]. Warm-ups and exercises can be chosen to address specific challenges in the repertoire to be practised. Within the repertoire, research suggests that students primarily work on new pieces without having strategies to review former works [Jorgensen, 1998]. At the conservatoire level, practice sessions are often started without setting clear goals for the time period [Jorgensen, 1998]. While related to activity selection, specific goals provide focus to sections of time devoted to a particular exercise or piece of repertoire. These goals might be technical, focusing on developing a particular skill or executing a specific task, or musical, focusing on the achievement of a particular artistic aim that may require technical experimentation to achieve. In either case, the goal should define not what is to be practiced but why. Furthermore, there is no standard answer as to how much time should be spent in practice. Practice time and frequency tends to increase with experience [Sloboda and Davidson, 1996], and surges in practice time tend to take place preceding examinations [Hallam, 2001] and both before and after lessons [Lehmann and Ericsson, 1997]. As excessive practice can trigger injury and exhaustion, it stands to reason that efficient practice should be the goal. Musicians should not fill a predetermined amount of practice time; rather they should spend what time is needed to achieve specific goals as part of a purposefully planned schedule. Practice time is also limited by practical constraints, so taking into account one's daily schedule and availability of practice space or equipment are key. Once the activities, goals, and time frame of a practice session have been planned, the musician must still employ appropriate rehearsal strategies and techniques to achieve the planned outcome. Jørgensen, [Jørgensen, 2004], described four common strategies used in musical rehearsal:

- Mental rehearsal: "nonplaying" practice in which musicians employ imagery, score reading, and audiation to reflect upon the music to be played, proving physical rest while developing a mental model of the intended outcome to which the physical performance can be compared. Transitions from mental rehearsal to physical rehearsal can be fluid.

- Rehearsing challenging parts: in their practice, musicians will identify exercises, techniques, or pieces of repertoire that they find particularly challenging. Common strategies to address these issues include repetition, incorporating technical etudes into practice that address the particular issue in question, breaking the section into component parts, and slow practice.

- Rehearsing the whole piece versus smaller parts: musicians may choose to practise exercises and repertoire from beginning to end without pause, play through while stopping to address particular issues, or concentrate on individual parts of the score. Each offers advantages; rehearsing the whole work allows for sections to be understood in the context of the whole and develops fluid transitions between components; rehearsing parts can allow for the efficient address of challenging sections, improve memorization [Chaffin et al., 2010], and result in technically stronger performances [Williamon and Valentine, 2002]. Thus, musicians should employ a combination of these techniques. The choice of where sections begin and end differs between pieces and individuals, often driven by musical and structural features of the work.

- Fast versus slow rehearsal: While slow practice is a common strategy in approaching unfamiliar works and ensuring accuracy in those well known, executing a work at a different speed to the intended final performance can activate a different set of physical movements[Winold and Thelen, 1994]. This must be weighed against the advantages of having more time to perceive and classify mistakes, thus an alternation between the two approaches is usually recommended.

6.2.1.2 Evaluation and review

A core element of TELMI project was the development of a system's aural, visual and motion feedback components will focus on review and evaluation of playing. As [Jørgensen, 2004] found in one study that 79% of conservatory students did not regularly review the results of their practice, for TELMI providing this feedback was a core element, finding a way to encourage new evaluative behaviours. Part of this incentive came from data collected being used to drive planning and execution, as described above, but it also required systems that provided engaging and intuitive feedback that could be quickly understood and acted upon. Speaking about multisensory feedback, musicians tend to compare their practice to internal and external models of performance. These may be mental models of how a technique should sound or feel, or audio and video recordings of peers, teachers, and masters performing the same works. Experienced performers are expected to transition from imitating the performances of others to developing their own models for performance, though research highlights how even experts can benefit from this practice [Tânia et al., 2005]. Finally, the performance is fundamentally different from practice, usually involving new (and larger) spaces with different acoustic responses, the presence audiences and/or evaluators, different attire, a higher degree of risk associated with errors and memory lapses, and higher levels of arousal and anxiety. Musicians should practice for these situations, learning how they react mentally and physically to the demands of live performance and developing appropriate strategies to counteract them.

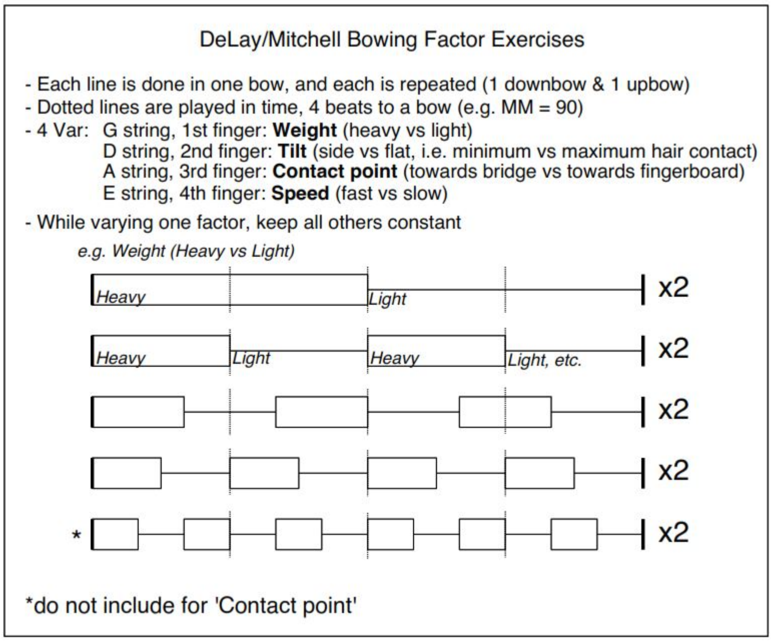
6.3 Definition of a multimodal archive for recordings of expert models of success

For each skill highlighted in the previous section, the TELMI project specified musical exercises chosen by traditional violin methods or specifically developed for its practice and assessment. Those skills comprised the following exercises:

- Handling the instrument
  - Posture
  - Holding the violin
  - Holding the bow
- Techniques of the right hand
  - Bow control
  - Bow direction
  - Bow changes
  - Bow tilt
  - Bow speed
  - Bow weight
  - Bow contact point
  - String crossing (Kreutzer, Etude No. 13)
- Techniques of the left hand
  - Articulation
  - Intonation (Schradieck, School of Violin Technics, Book 1)
  - Shifts (Yost, Exercises for Changes of Position)
  - Double Stops ()
  - Vibrato
  - Trill (Kreutzer, Etude No. 14)
- Articulation
  - Détaché
  - Legato
  - Martelé (Kreutzer, Etude No. 7)
  - Pizzicato
  - Sautillé (Ševčík, Op. 3, Var. 16)
  - Spiccato (Ševčík, Op. 3, Var. 34)
  - Staccato (Kreutzer, Etude No. 4)
  - Ricochet
- Scales and arpeggios (from ABRSM, Galamian and Flesch)

To this end, exercises have been selected that directly or indirectly address these techniques. These were selected under the guidance of TELMI RCM co-investigator Madeleine Mitchell and supported through consultation with performing, teaching, and student violinists. The exercises take three forms: (1) those sourced from the standard published catalogue of exercises, including those of Schradieck, Ševčík, and Kreutzer, which the TELMI survey found to be the most popular; (2) those sourced or adapted from the Associated Board of the Royal Schools of Music (ABRSM) examination syllabus; and (3) customised exercises developed by Madeleine Mitchell to address specific techniques with specific focus on the capabilities offered by non-notated feedback (e.g. the bowing exercises). The use of both custom and pre-existing exercises was deliberate. The bespoke pieces are intended to capitalise on the unique possibilities offered by audio, video, and motion-capture feedback, while the published sources draw on centuries of pedagogical insight and tradition to give students and teachers a common frame of reference when approaching the TELMI system. The fundamental techniques required by violin technique cannot easily be isolated. For example, one cannot practice bow speed without making some choice about the weight or tilt of the bow, nor can one practise articulation without handling the violin. Thus, while many of the following exercises were chosen to directly address a particular technique (e.g. Kreutzer Etude No. 7 for the Martelé articulation), fundamental techniques such as posture and handling the instrument are reflected and re-articulated within any of the exercises and thus do not require a specific test. As posture and use of the body is a highly idiosyncratic technique, depending greatly on the physiology of the individual player, the student will have access to TELMI's feedback systems and, with the guidance of a teacher, be able to develop a self-awareness of their body use throughout the various exercises. Regarding bowing, general control of the bow is a reflection of the underlying principles of tilt, weight, speed, and contact point of the bow. Thus, the custom bow factor exercises reflected these four conditions, and along with the 'slow bow' exercise allowed not only for pseudo-isolation of these techniques but also careful consideration of the bow change and direction. Regarding the bow change, as the fundamentals of this technique will vary based on the context in which it has been used, and this context will appear in the various exercises below, a single exercise was not chosen. An 'open' version of the TELMI feedback system, in which students are not following a specific exercise but rather given real-time feedback on their acoustic and physical actions, will also allow them to experiment with features such as the bow change under guidance of the teacher. The six techniques of the left hand (articulation, intonation, shifts, double stops, trills, and vibrato) are addressed with specific existing and custom exercises as listed below. The same is true for the list of articulations (to which the Sautillé has been added) through specific exercises. Furthermore, Madeleine Mitchell refined the following set of bow factor exercises following initial development by her teacher Dorothy DeLay at The Julliard School. These exercises isolated the four fundamental bowing techniques of weight, tilt, contact point, and speed by asking the violinist to vary one while holding the other three constant. This not only allowed the musician to develop control of each physical technique, but also to explore the acoustic result of each. Fig. 6.3 below represented how the exercise is to be performed. Each line represents a single bow stroke, divided into four beats. Within the stroke, the violinist alternates between two extremes of the technique (e.g. heavy and light, for weight), first separating the bow into 2 sections, then 4, 6, 8, and 12. The degree to which the violinist varies the two conditions is at their and their

Figure 6.3: Graphic notation for the custom DeLay/Mitchell Bowing Factor Exercises.



teacher's discretion – the focus should not be on the degree of difference produced but rather the consistency of tone between corresponding sections, and that an immediate difference can be perceived. Each is repeated; one upbow and one down-bow. Though the left hand is not explicitly involved in each exercise, each is performed using a different finger on a different string, resulting in a variety of held pitches and left hand positions while completing the exercises. As these are non-notated exercises, TELMI users have being presented with demonstrative videos describing and demonstrating their correct performance. They will also be able to follow along a graphic notation similar to the one in Fig. 6.3, where an indicator moves across the graphic showing their place in the exercise and, where possible, giving real-time feedback on the performance.

6.4 Technology in learning instrument

The traditional and often restrictive master-apprentice model of instrumental teaching, not to mention the artistic and humanistic nature of music-making, can give the impression of a domain incompatible with technological innovation [Gaunt et al., 2012]. As will be described, technology use is prevalent (though still underused) in the general music classroom and as a driver of research in prototype learning systems (though evidence of use of those systems in general practice is scant). Within instrumental learning, technology is generally used to access online resources, for distance learning and communication, for schedule organization, and in traditional roles (e.g. audio recorders, tuners, metronomes, etc.). While published rates of technology use among instrumental teachers are not available, this section will close by summarizing ongoing TELMI research that seeks to address this gap in the literature. Within the field of education, technology has been seen as a tool to transform, modernize, and generally improve the classroom [Purves, 2012]. The music classroom has followed suit. The suite of electronic keyboards and Apple computers has become ubiquitous within primary and secondary schools, along with professional-grade recording equipment and software packages allowing each student to compose, record, edit, and play back their creations. Thus, the role that technology plays, and is expected to play, is ever widening. Himonides and Purves, [Himonides and Purves, 2010], survey these roles as follows, identifying how music technology is expected to allow users to:

- "Become better musicians";
- Understand music and/or the wider impact that music has on our lives and ongoing development;
- Record, capture, experience, study, create, compose, document, analyse and archive sound and music;
- Enhance our lives through experiencing music in new ways;
- Facilitate the communication of our music (performances or compositions);
- Provide wider access to other people's music (individuals' as well as other cultures' in general);
- Provide access to music for people with special needs and requirements;
- Monitor and assess our teaching practices in the music classroom;
- Monitor and assess our students' development and learning experiences;
- Research, scrutinise, assess and evaluate current educational theories (and their applications to practice) and allow the development of new theories, practice and policy for music education."

The majority of these benefits can be accomplished by technologies that have now been commercially available for decades, including portable audio- and video-recording devices as well as desktop computers. While best practices for effective use have not been formalized, by the mid-2000s teachers were independently developing strategies to incorporate recording and edited playback, videoconferencing demonstrations and collaborations, and web-based instructional videos into their teaching [Anderson and Winawer, 2005], and calls for more professional development were made following the demonstrated benefits of music technology-based teacher workshops [Bauer and Dunn, 2003]. By 2003 the UK government reported that 24 percent of secondary classrooms made substantial use of technology and as a result, 30 percent of secondary teachers reported a substantial positive effect on teaching [Jones, 2006]. In 2010 an independent review on Music Education in England for the Department for Education recommended that:

> "further work should be undertaken to develop a national plan for the use of technology in the delivery of Music Education – and to ensure that the workforce is up-to-date with the latest developments. This review should examine how technology could enable better teaching of music (particularly in rural communities) as well as ways in which new methods of creating music that embrace technological innovation are taught in the classroom" ([Henley, 2011], p.30)

Furthermore, recent studies have observed that, despite ever increasing portability and affordability, technology remains underused [Fautley, 2013] with the greatest barriers being a lack of availability, technical competence, and staff support [Gall, 2013]. This is analogue with data from the European Commission, [EUROPEA, 2013], which showed a substantial increase in numbers of computers and quality of broadband access in European schools from 2006-2012 and marginal growth in use, though fewer than half of teachers were making use of ICT in more than 25% of their classes. Where technology is being used in the music classroom, it is taking advantage not only of increased school resources but also the explosion in mobile technologies to deliver enhanced instruction. This includes training in music and rhythmic reading, aural skills, composition, and basic interaction with instruments including the guitar and percussion, though its most common use is still for administrative rather than curricular tasks [Webster, 2012]. The technological augmentation and digitization of instruments is also prevalent, not only in increasing the functionality of standard models such as the prevalent MIDI-enhanced keyboards, guitars, and percussion devices, but also in newly designed instruments that attempt to lower the technical bar for use and allow users to express a greater degree of creativity. These instruments may be hardware- or software-based, and have been used particularly with children [Lagerlöf et al., 2014]. Quantitative research on the prevalence of and attitudes toward technology use in one-to-one instrumental teaching is currently not available in the literature to the extent is has been documented in the music education domain. Existing evidence tends to be anecdotal, incomplete, or out-of-date, such as one question in a survey of 100 instrumental teachers by [Barry and McArthur, 1994] who found that the majority of instrumental music teachers did not use or encourage their students to use software-based music learning tools, though the technologies available would have been limited at the time. Outside the music classroom, the Internet plays an ever-increasing role in expanding access to music-learning resources. On video sharing platforms such as YouTube, a plethora of instructional videos exist for learning to perform virtually any musical instrument which are used not only by individuals but being incorporated into classroom curricula [Waldron, 2013]. As of December 2019 a search of YouTube with the phrase "learn violin" produces over 420.000 results. These videos tend to feature individual musicians demonstrating specific techniques, challenges, exercises, and pieces of repertoire through a combination of verbal instruction, performance modelling on their own instrument, or through a mock lesson with a student. Focus tends towards beginners, often in a structured series of lessons; a popular example is The Online Piano and Violin Tutor at www.youtube.com/user/theonlinepianotutor, with over 225,000 subscribers as of July 2016. Videos are also available for more advanced players, incorporating in-depth technical, philosophical, and aesthetic discussions of musical challenges, including issues of effective practice and career challenges as a violinist (see, for example, "The Art and Science of Violin Vibrato" at www.youtube.com/channel/UCzkT7lUQcMUMawwTlkdNZ3Q). While such videos offer advantages of accessibility, choice, and low economic cost (provided the musician has access to a computer and broadband connection), with the crucial feature of allowing sections to be paused, slowed, and reviewed at the user's discretion, they offer inherently one-directional communication of information. Users cannot change their visual perspective while watching, have feedback on their own playing, or ask clarifying questions through the video alone. Other online resources include websites devoted to violin learning. Some offer violin-specific training, variously combing videos, exercises, visual and written guides, and user forums, often advertising increased opportunities to contact expert violinists via message or videoconference, occasionally through formal lesson subscriptions (e.g. www.alisonsparrow.com; www.dakapp.com; www.fiddlerman.com; www.violinfromscratch.com; www.violinonline.com; www.violinlab.com). While these sites offer the added benefits of potential electronic communication with experts and a wider variety of electronic resources than can be offered through a platform such as YouTube, the same inherent issues exist of a lack of individualized, real-time feedback systems due to primarily one-directional communication channels. Beginners can supplement this information by visiting sites offering lessons in basic concepts of musical and rhythmic notation, music theory, and ear training (e.g. www.daveconservatoire.org; www.musictheory.net; www.mymusictheory.com; www.smartmusic.com). Research suggests that music instrument learners tend not to take advantage of opportunities to communicate with experts through these online channels. In one study [Thorgersen and Zandén, 2014], nine students aged 20-30 and beginning their teacher training were asked to learn to play a musical instrument new to them using only resources available online. Given the freedom to use any tools available, the majority of resources came in the form of instructional videos, charts, and resources. Apps, forums, and communication tools allowing for a two-way transfer of information with peers or experts or some degree of individualised feedback were generally not employed. Ever-increasing hardware and software platforms for real-time videoconferencing are opening new avenues for music instruction. Free platforms such as Skype are seeing increased use in private and conservatoire-level music education, and while they have been found to preserve the basic "feel" of teacher-student interactions and offer obvious advantages in increasing access to instruction, evaluation, and audition by reducing travel restrictions, issues of low audio and video quality, high latency, hardware issues, lack of physical contact between teacher and student, and difficulties in maintaining eye contact have been found to limit efficacy [Kruse et al., 2013]. Some lower-level skills seem to be well suited to such instruction; one study found that students learning sight-reading skills via a videoconferencing lesson, in conjunction with a pair of MIDI keyboards that were able to transmit performance data, were able to improve their skills as effectively as those receiving a traditional face-to-face lesson [Pike and Shoemaker, 2013]. While advanced systems are being developed that aim to maximize audio and video quality and minimize latency (e.g. Polycom; LOLA) the cost and expertise required operate such systems restricts their use to the organizations who can afford them. Finally, numerous technologies have been in common use by musicians for decades, transitioning from standalone devices to integration within mobile devices in recent years.

- Metronomes
- Digital tuners
- Audio/video recorders: students are commonly advised to record their practice sessions and performances to enhance opportunities for review and reflection. Video recording in particular allows for musicians to see movements that are otherwise impossible to perceive visually. Recent research has demonstrated significant differences between self-assessments of live versus recorded versions of the same performance, demonstrating the differences in perception and perspective reviewing one's performance can bring [Silveira and Gavin, 2016].

6.5 Prototype systems of violin learning

Numerous systems have been developed to enhance feedback in instrumental control and learning, though there is not clear evidence demonstrating common use among musician learners and teachers nor the attitudes or barriers to increased use. The current section offers a review of those developed for the violin (or similar stringed applications), though systems based on similar principles of sound, visual, and motion analysis can be found for other instruments with adaptation to their particular idiosyncratic features. The i-Maestro system (Interactive Multimedia Environment for Technology Enhanced Music Education and Creative Collaborative Composition and performance, [Ng and Nesi, 2008]) was developed with funding from the European Commission Sixth Framework Programme. It comprised a suite of tools that included a 3D motion capture system (termed the "Augmented Mirror") based on a 12-camera system and small spheres worn by the user, from which information regarding bow motion and posture could be extracted and synchronized with audio and video data. While early tests with users were positive, the high degree of cost and technical knowledge in operating the system restricted mass use. Other features included a music exercise generator, and tools for enhancing distance learning. Hall and O'Donnell, [Hall and O'Donnell, 2011], developed a computer-based system for the improvement of violin bow control and the choice of bow direction, speed, and changes. It read and analyze a printed musical score and, after comparing it to a model of bowing motions and their effects, determined which combination of strokes was more appropriate. The student was then given a prediction of the ideal position and speed of the bow for each stoke, with potential difficulties highlighted, so that the student may decide whether to modify his gesture choices accordingly. The student also saw a textual representation of these recommendations in the score. The system aimed to determine and help solve problems in bowing, especially for students in the early stages of their learning. In the same category of gesture-focused approaches, Johnson and colleagues, [Van Der Linden et al., 2009], introduced an inertial motion capture system to help teach bowing technique and correct posture in novice students. It provides both visual and vibrotactile feedback, guiding the movements of the student along the correct trajectory. Other systems have focused on audio analysis. The 'Digital Violin Tutor' (DVT) by Yin, Wang, and Hsu, [Yin et al., 2005], is an integrated system for early violin students that combines audio transcription of violin performance with computed visualisations of the playing motions, notes, and/or fingering positions on the neck of the instrument where the notes are to be played. The entire system has been implemented with off-the-shelf hardware, as it was destined to be practical in home environments. A similar system is the 'Music-driven Digital Violinist' (MDV) by Yin et al., [Yin et al., 2004]. This system automatically generates an animation of the violinist based on the music performed, thus introducing visual feedback on top of audio recordings. The system aims to act as an interactive demonstrator for beginners by allowing them to interactively view correct playing motions from different viewing angles. 'Interactive Digital Violin Tutor' (iDTV) by Lu et al., [Lu et al., 2008], is another system that combines audio and video feedback. This system is also destined for a household acoustic environment and focuses on pitched non-percussive (PNP) sounds, such as those from the violin. The system employs two ordinary webcams and one microphone as peripherals, recording the audio and a frontal and top view of the performer. In the same category, the 'Tuning Perception Test' (TPT) by Hopkins, [Hopkins, 2014], measures orchestral string players' instrument tuning skills. Preliminary data offered by the system suggested that the pitch perception skills necessary to tune a stringed instrument are gradually refined over the course of learning the instrument. Other systems focus on finger placement. While not violin-specific, the 'guitAR' augmented reality application by Löchtefeld et al., [Löchtefeld et al., 2011], assists students playing a string-based instrument with the use of a portable projector attached to the instrument. Instructions are projected directly onto the strings and frets of a guitar so that the user may easily see where the fingers should be placed in order to play a certain chord or a tone sequence correctly. This approach enables the student to move their guitar freely and keep their attention focused on the instrument. Other systems measure finger placement directly; one by Grosshauser and Tröster, [Grosshauser and Tröster, 2014], uses force-sensitive-resistors (FSRs) built into the fingerboard of the instrument, measuring pressure of the individual fingers and giving feedback on their placement. Finally, some systems focus on expressivity and musicality. Ho et al., [Ho et al., 2015], developed VQVA-Sys, a computer-based visualised learning system for playing violin vibrato (a technique of oscillating the pitch of a held note to make it more expressive). The use and treatment of vibrato varies greatly between musicians, regional and historical styles, and personal tastes, thus violinists are expected to be able have control of a wide range of vibrato techniques at their disposal. After having used the VQVA-Sys system, most of the students were able to rapidly adjust their vibrato technique by simply focusing on the motion used to produce the sound. The 'Feel-ME' (Feedbacklearning of musical expressivity) project is an ongoing computer application by Karlsson, Liljeström, and Juslin, [Karlsson et al., 2009], aimed at computationally representing musical expression by modelling the slight variations in timing, volume, and pitch that add character to a performance. Their pilot research found that a computer program may improve performers' abilities to express emotions through their performance, though they also found that performers are reluctant to embrace this novel technology.

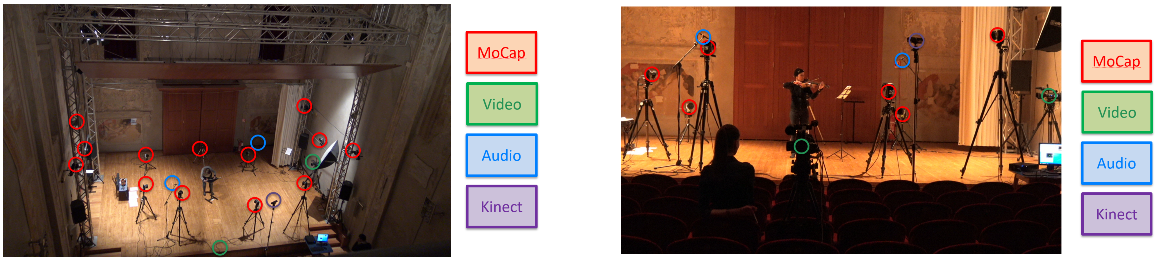
6.6 Multimodal archive of recordings

To create a model of expert in violin playing, as a consortium, we organized recordings of high-level professional violinists from Royal College of Music of London. This dataset and the recordings' protocol followed, was presented in the paper [Volpe et al., 2017]. We recorded four internationally renowned professional violin players for the corpus of exercises presented above. Recordings consisted of motion capture data (performer, violin, and bow), instrument and ambient audio, video (frontal and lateral view), and data from physiological sensors (EMG as captured by Myo sensors). Equipment included:

- A 13-cameras Qualysis motion capture system;
- Two video cameras JVC GY-HD251 (720p, 50 fps);
- One Kinect v.2, capturing both live video and depth-map;
- One pick-up Fishman PRO-V20-0VI connected via radio (AKG UHF wireless system PT40-SR40) mounted on the violin;
- Two Neumann KM184 microphones for ambient audio;
- Two Myo sensor for EMG data.

Fig. 6.4 shows a top and a frontal view of the environment in which the recordings took place and shows where the equipment was located. Recordings were carried out at Casa Paganini-InfoMus research Centre of DIBRIS, University of Genova.

The violin was endowed with 6 small lightweight reference reflective markers and with a pick-up microphone. A 6DOF rigid body was defined to track the violin in the local systems of coordinates. Virtual markers were used for the violin strings. These are markers that are placed at each string-end during the calibration phase. Then markers are removed for the performance, as they would be very intrusive. During the actual tracking, virtual markers are reconstructed from the reference markers. After the first recording session with Madeleine Mitchell, the same



(a) Top view (b) Frontal view

Figure 6.4: Top (a) and frontal (b) view of the set-up. Colors show the locations where the video cameras of the motion capture system, the video cameras for video recordings, the microphones for ambient audio recordings, and the Kinect sensor were positioned.

violin was used for the next three performers. This choice had a twofold benefit: (i) it reduced the variability introduced by using different music instruments (i.e., it reduced the number of variables to take into account), and (ii) it enabled speeding-up the recordings, since the violin could be prepared in advance by putting motion capture markers and the pick-up microphone in appropriate positions on its surface. Some time was needed for the players to get acquainted with the new instrument. The players, however, are all professionals. Fig. 6.5 shows where the reflective markers, the virtual markers, and the pick-up microphone were positioned on the violin.

The bow was also endowed with lightweight reflective markers and a 6DOF rigid body was defined to track it (see Fig. 6.6) The same bow (prepared in advance) was used with two of the four players. Both the violin and the bow were kindly provided by Alberto Giordano an internationally renowned luthier, conservator of the Cannone, the violin owned by Niccolò Paganini. Markers on the body of the player were placed as follows: right and left front head, right and left back head, head top, C7 spine, spine fifth thoracic vertebra, spine tenth thoracic vertebra, left and right shoulder, left and right scapula, left and right knee (inner and outer side), right and left pelvic bone back, sacrum bone, right and left toe, right and left elbow (inner and outer side), right and left wrist (inner and outer side), right and left palm index finger, right and left palm pinky finger. A motion capture suit of appropriate size was used. Moreover, two Myo bracelet were placed on the left and right forearms.

We recorded the data synchronizing them through EyesWeb XMI platform, [Camurri et al., 2005, Volpe et al., 2016]. Synchronization is based on SMPTE time-stamps. Kinect and MYO data were synchronized by EyesWeb XMI using proprietary time-stamps. We informed in advance the violinists about the requirements for the recording, e.g. the need to wear a motion capture suit and the physiological sensors as well as the need to play a violin prepared with markers. Upon arrival at the location of the recordings, after initial welcoming and further explanation of the purpose and procedure of the recordings, each participant was dressed with a motion capture suit of appropriate size and was given time to get used to play in the required conditions. At the

Figure 6.5: Reflective markers and a pick up microphone are placed on the surface of the violin. Moreover, virtual markers are used.

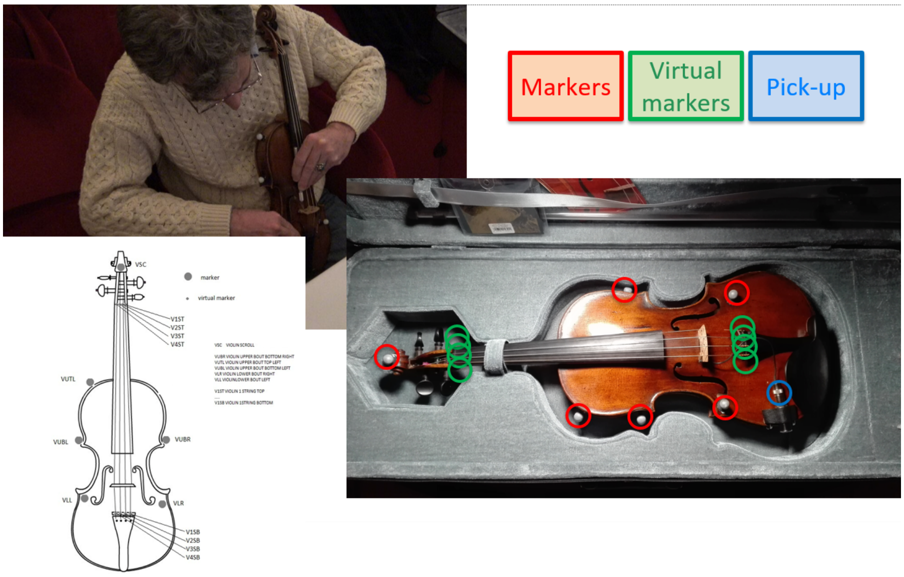


Figure 6.6: Positions of the reflective markers on the bow

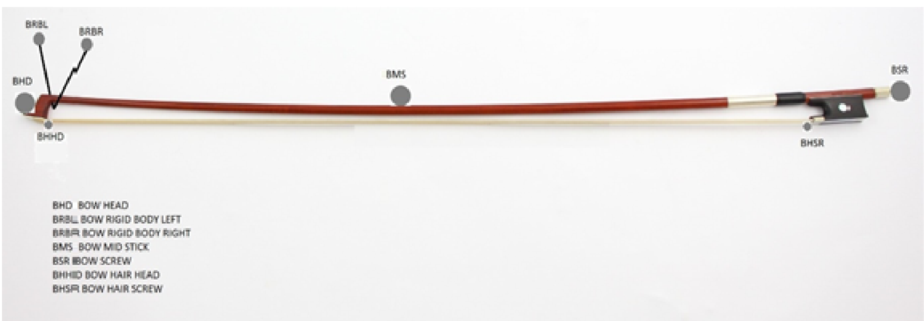
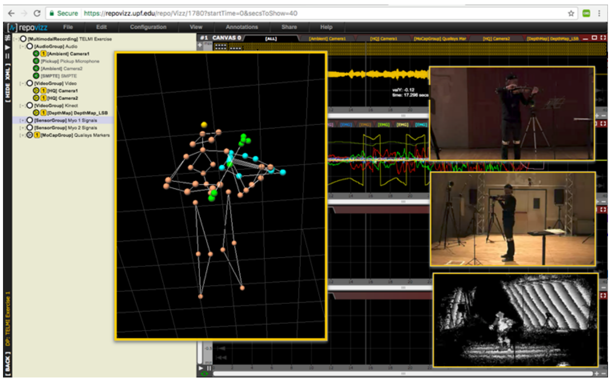


Figure 6.7: A sample item of the corpus in the repoVizz web interface



same time or immediately before starting the recording session, the motion capture system was calibrated and a recording test was performed to check whether all the required equipment was working properly. During the recording session, the player was free to play the exercises in the order s/he preferred. The player could replay each exercise how many times s/he wanted, until s/he was satisfied with the performance. Players were also given the opportunity to check the recordings after the session and to record again the exercises they were not satisfied with. The players were informed of the purpose of the recording and signed a consent form.

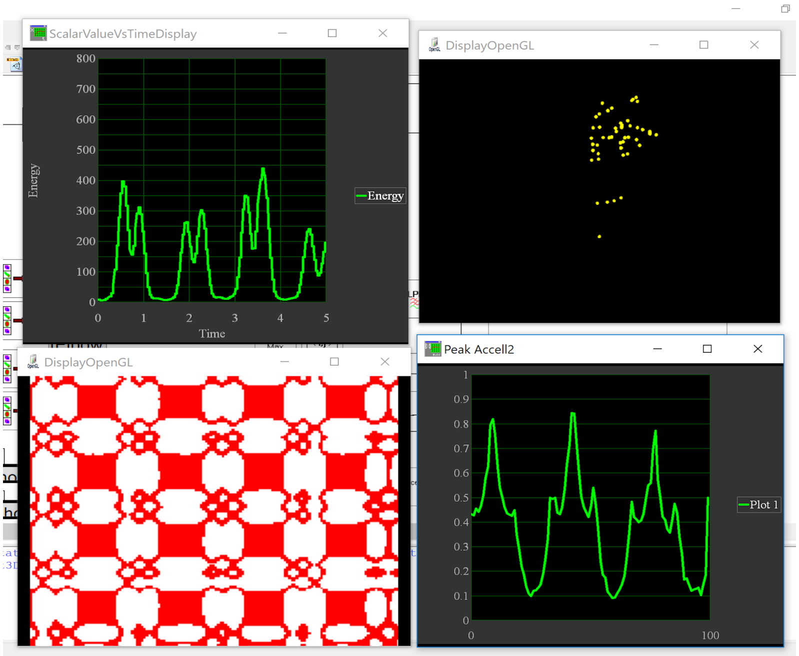
After these recordings, in the following months of 2017 and 2018, we recorded the same pedagogical materials also with 4 students, with different grade of study and expertise. The recorded material was post-processed and uploaded in the repoVizz repository, where it is publicly available. Fig. 6.7 shows a sample item of the corpus in the repoVizz web interface. EyesWeb Mobile interfaces were also developed for preview and play back of downloaded data. Data has been then analyzed to extract audio and movement features. Concerning the latter, computed features include kinematics (velocity, acceleration, jerk, curvature of relevant trajectories such as hands and head, angles between bones, kinetic energy) and higher level features, such as openness, lightness, sway, tension, suddenness, coordination. In fig. 6.8 shows an application of Recurrence Quantification Analysis, [Marwan et al., 2007], to the kinetic energy of the right wrist. Recurrence rates of signals captured from players having a different level of skill can be compared, e.g., as a contribution for assessing the technical quality of the music performance.

The framework of the project was based on multimodal pedagogy mediated by the technology itself, comprising audio features, violin controls and body motion.

6.6.0.1 Audio features

Audio features have been extracted and analysed by project partner and coordinator of UPF, in order to characterize the properties and quality of the sound, both in the temporal and spectral

Figure 6.8: Recurrence Quantification Analysis applied to kinetic energy of the right wrist. In the bottom, recurrence plot and probabilities of recurrence of the signal in the top.



domains. Important temporal features are segmentation of notes into segments attack, sustain, and release; the duration of each of these intra-note segments; the intensity or slope of the note attacks; intensity and note dynamics; temporal envelopes of energy and pitch as well as lower level features such as zero crossing rate, and so on. Important spectral features of hte sound are harmonic content; the stochastic content; inharmonicity; spectral envelope features such as MFCC, slope, kurtosis, skewness and so on.

6.6.0.2 Violin Controls

Instrument controls are physical actions articulated by the musician that have direct impact on the properties of the sound. The main controls in violin performances can be divided into lefthand controls, which directly affect the pitch and will be therefore analyzed as an audio descriptor, and bowing controls that will be measured with the systems presented above. The most important bowing controls are depicted in the figure below and are the played string, the bow transversal position, its derivatives velocity and acceleration, the bowing angles skewness and tilt, the bow to bridge distance (bbd) and the bow width in contact with the string. In next pararaph, we present some results obtain in order to facilitate violin controls.

6.6.0.3 Human Body Motion

The main descriptors related to posture and motion are the position, velocity and acceleration of the joints and bones of the body, and the angles between bones together with their respective angular velocities and accelerations. Further descriptors concern quality of the movement performance including for example, energy of a movement stroke, whether it is sudden or sustained, whether it is direct or flexible, whether it is heavy or light, whether it is determined or hesitating, its fluidity, its impulsiveness, and so on. Such descriptors come from previous studies on human movement analysis and, specifically on analysis of movement in music performance. The interface we developed for TELMI, the SkyMotion system, allowed to control some body parameters, with a Kinect-based software. Using automatic evaluation, the student needs to be encouraged to not achieve single static postures, but instead develop an awareness of the body and positions that allow for free movement to achieve their desired outcome with minimal tension. Four sections need to be highlighted:

- head and neck;
- left arm and shoulder;
- right arm and shoulder;
- body, including general posture and weight distribution.

Low-level features from motion capture systems include position, velocity, acceleration, jerk, local curvature, and smoothness of the tracked 3D points, head direction, kinetic energy of the cloud of tracked points. Low-level features from Kinect include position, velocity, and curvature referring to the trajectories of the 3D points of the Kinect skeleton, head and overall movement direction, overall kinetic energy of the cloud of reliably tracked points, or a quantification of the overall amount of movement (e.g., the Motion Index) from the 2D blob Kinect retrieves, in case the skeleton is not reliable. Fig. 6.9 shows an interface developed for EyesWeb Mobile to control the low-level feature computation application (patch) running in the EyesWeb platform.

The patch computes the low-level features and can either store them for off-line analysis or send them to the ViolinRT library by UPF for real-time visualization. This enables providing the student with either fast real-time feedback or more accurate off-line analysis of his/her performance. Regarding low-level postural features (from both motion capture and Kinect), we developed some EyesWeb XMI patches for computing the following ones: (a) openness (i.e., contraction / expansion) and (b) leaning of the posture on the frontal (right / left) and sagittal (back / forward) plane. Tension is computed from the angles between the adjacent lines identifying feet (the line connecting the barycentre of each foot), hip, trunk, shoulders, and head directions, and is inspired by classical paintings and sculptures where such angles are exploited to express postural tension. In TELMI tension was re-conceived to adapt it to the special case of

Figure 6.9: **An EyesWeb Mobile interface to control extraction of low-level features in EyesWeb XMI.** On the bottom left, the body joint and the feature to be extracted on its trajectory can be selected (e.g., speed of the right wrist). On the bottom left, the user can select to extract the direction of different parts of the body (e.g., head, shoulder, trunk). On the top, the cloud of the tracked points and the graph of the selected feature are displayed.

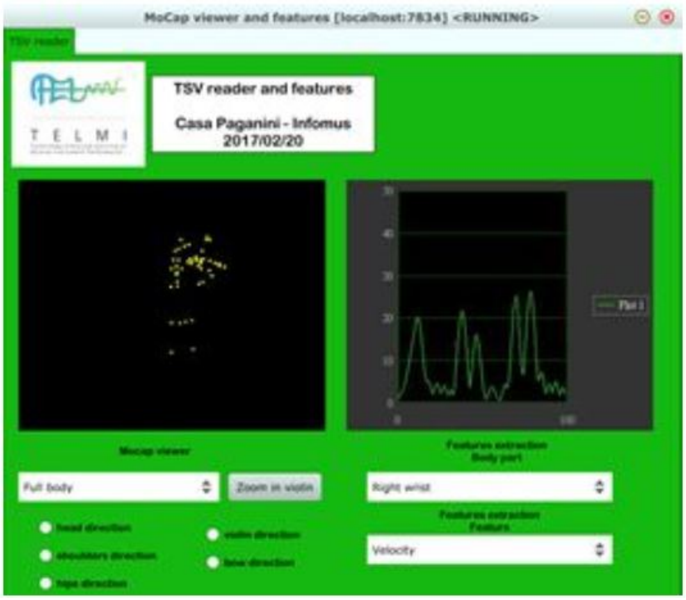
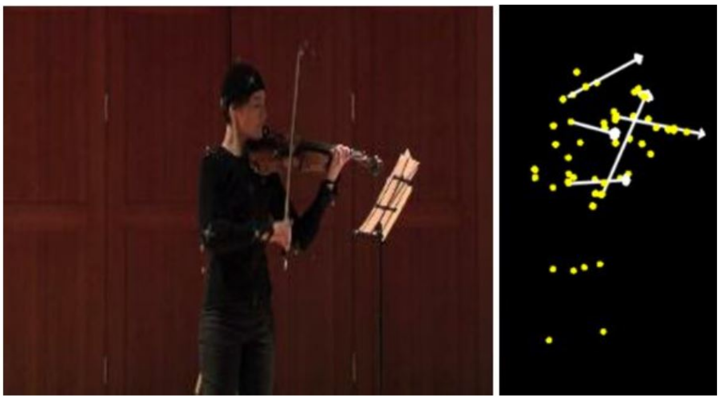


Figure 6.10: **An example of analysis of tension**: direction of the head, trunk, shoulders, violin and bow are computed and their differences analysed. Tension, as a postural feature, is inspired by classical paintings and sculptures where such angles are exploited to express postural tension.



violin performance and to take into account also the directions of the violin and of the bow. Fig. 6.10 shows the directions relevant for tension extracted from an exercise included in the TELMI reference archive and played by professional violinist Eulalie Charland. After discussion with violin teachers and players, tension seemed worth of a relevance as an overall postural descriptor for violin performances.

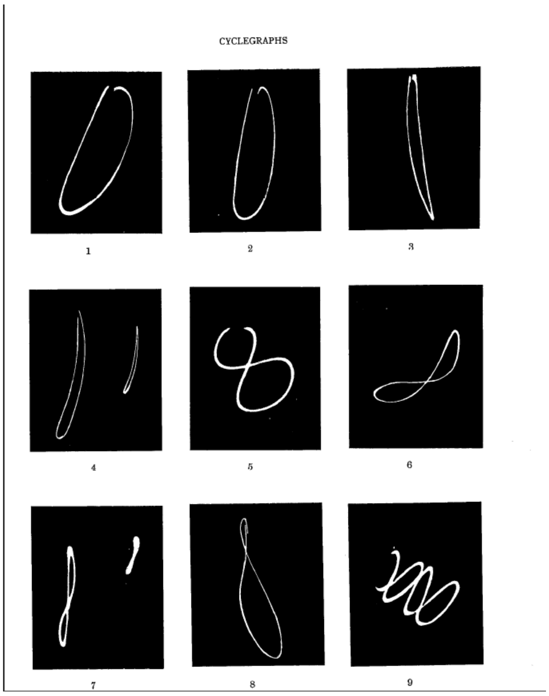
Concerning the mid-level features, we developed EyesWeb patches to compute the following ones: dynamic of postural openness, suddenness, and sway. Since regularity (intended e.g., as a low variance in the motion) appears to have a special role in professional violin playing, as it emerged from the analysis of the TELMI reference archive of professional performances as well as from discussion with teachers and professional players, a specific research activity started on the analysis of coordination, intended as intra-personal synchronization as a component of regularity. Initial steps include application of Recurrence Quantification Analysis [Marwan et al., 2007] to relevant movement signals (e.g., the kinetic energy of the right wrist) to (i) identify possible recurrences in the signals and (ii) compute synchronization of body parts and movement signals (e.g., synchronization with leaning or with body sway).

6.7 Skill Performance Metrics for Automatic Evaluation

6.7.1 Articulation of bowing trajectories

In 2017 we approached the multimodal archive previously recorded, considering one of the most challenging element of violin learning: bow articulation. Methods introducing violin pedagogy commonly begin illustrating how to hold properly violin and bow, as well as the proper posture

Figure 6.11: Hodgson's cyclographs



to support body movements without impeding bow arm [Galamian, 1962]. A great importance in the violin pedagogy is given to techniques of the right hand that is responsible for most of the sound produced.

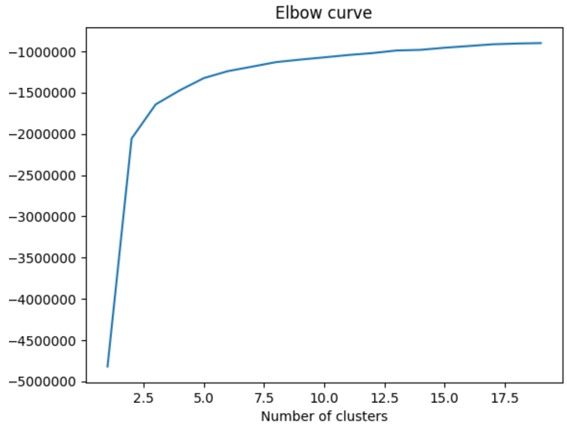
Several studies investigated the movement of the bow, finding a strict relationship between motion characteristics and quality of the performance, e.g., the bowing motion should be fluid [Galamian, 1962] and circular [Starker and Bekefi, 1965]. One of the first and most influential technical approaches to the study of bow movement is Hodgson's *Motion Study and Violin Bowing*, published in 1934. In his famous work, Hodgson used early methods of photographic motion tracking to study the circular nature of bowing technique in cyclographs (see Fig. 6.11). The controversial insights of Hodgson's work, showing that the bow's trajectory is always curved, has caused an animated pedagogical debate, but the knowledge of the curved nature of bowing has influenced the pedagogy of the last century, giving to violinists an explanation and a metaphor to understand the correctness of their movements, since it is generally not common for students or teachers to see their own playing movements represented in a visual way. The body of scientific research related to music education has grown significantly in recent decades, e.g., see [Rauscher et al., 1997], as well as the trend in developing sensors-based systems to use bow gestures in interactive performance [Machover, 1992], [Nichols, 2002], [Overholt, 2005]. Despite this growth, since Hodgson's work, technology has rarely been applied to music pedagogy and usually restricted to other domains, such as to audio and video recording and playing. The aim of this preliminary work was to explore whether multimodal systems and machine learning techniques can be used for analyzing bow trajectories as a means of contributing to music performance pedagogy, by working on selected recordings of renowned performers and teachers recruited by the Royal College of Music in London. Bow control is a central musician's skill, giving the violinist the ability to direct bow's motion during playing. In his work, Hodgson divided bowing movements in three categories of motion:

- Movements across the string, influenced by tilt, speed and contact point techniques;
- Rotation movement around the string, that allows changing across strings and changing in direction;
- Movement towards and away from string, varying the weight and that is responsible for particular articulation effects.

For this work, we chose to focus our attention on articulation techniques, as one of the most delicate parts of violin education and as one of the elements that a multimodal system can help to analyze. From the entire TELMI archive, we selected the Martelé (from Kreutzer, Op.7), Spiccato, and Sautillé (from Ševčík, Op.3) exercises recorded by the four internationally renowned and esteemed professional violinists involved in the TELMI project. The choice of these exercises was made by considering the importance of these three different bowing techniques and the differences between them that are often confused and difficulty master by students. One of the difficulties of these studies was related to speed, because there is a common ground where one should be able to make the change from Spiccato to Sautillé and vice versa without changing any character of the sound profile. Furthermore, the mechanic of these two bowings are completely different. In Spiccato every single note is played actively, whereas in Sautillé the jumping activity is left quite exclusively to the resiliency of the stick. A further difference lies in the hand's motion: according to Hodgson, in fact, during Spiccato the bow designs an eight in the air and when Spiccato is quickened to Sautillé, the movement of the hand changes to an ellipse, but the bow continues to draw an eight in the air. Martelé represents, finally, a third type of fundamental movement of bowing, since it is at the basis of essential bowing techniques, such as Staccato, where the pressure is released between each stroke, and the bow speed has to be quite fast, yet light. We used EyesWeb to reproduce the recorded data and followed the bowing trajectories, trying to identify the Hodgson' bowing shapes. An example of such visualization can be seen in Fig. 6.14. After the recordings, data was segmented, by considering the musical structure, to isolate single bowing movements for each music phrase. We extracted 869 segments in total from 12 recordings we considered. Using EyesWeb we computed the 3D-trajectories of the tip of the bow, to check the presence of Hodgson's bowing shapes. We then computed six different features on each obtained segment, in particular:

- Acceleration

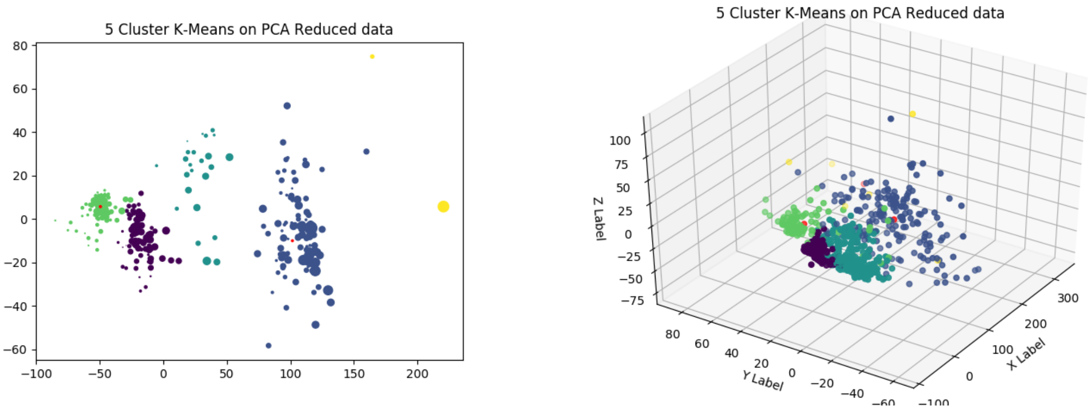
Figure 6.12: The Elbow curve. The red circle identifies the number of clusters (k =3) we choose for this study.



- Trajectory length
- Kinetic energy: an approximation of the overall energy spent while performing a movement with the bow. It is computed as the total amount of displacement in all the tracked points.
- Directness: this is a measure of the extent to which a given trajectory is direct or flexible. It is computed as the ratio between the Euclidean distance calculated between the starting and the ending point of the considered trajectory, and its length.
- Smoothness: this corresponds to the third derivative of the position and it has often been used as a descriptor to evaluate how a motion trajectory varies "slowly" over time [Flash and Hogan, 1985].

To obtain the same number of features for each segment, we extracted a cumulative histogram with 25 bins for each considered feature, resulting into a 150-dimension feature vector data. K-means was applied to all the feature vectors, to figure out whether the different kinds of articulation exercises we were studying can be distinguished from the characteristics of the bow motion. We then estimated the best number of clusters (i.e., the k parameter to seed the k-means algorithm), that is shown in the Elbow curve in Fig. 6.12.

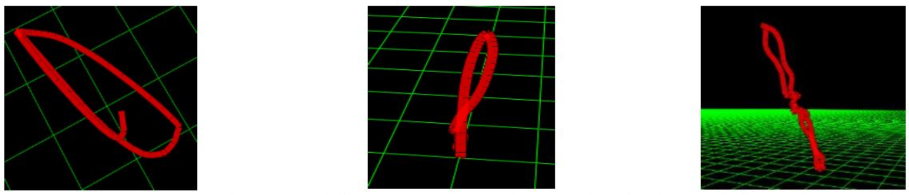
A value of k=3 was detected as the most appropriate value. Finally, we applied a PCA reduction to lower the dimension to the most representative 2 and 3 features, and visualized the clusters. Resulting data clusters are shown in Fig. 6.13 with different colors. As the figure shows, the considered segments were split mainly in three clusters. We verified a clear separation between three particular bow motion trajectories, i.e., segments belonging to the same cluster present similar trajectories.



(a) (b)

Figure 6.13: Three clusters obtained by applying K-Means on PCA reduced data

Figure 6.14: 3D visualization of the trajectories "drawn" by the bow while performing three different exercises. From left to right: a) Martelè: characterized by a circular trajectory, b) Sautillè: characterized by an "8"- trajectory, c) Spiccato: characterized by a lace trajectory.



We observed that most of the segments composing the purple cluster have been extracted from Sautillé performances, and segments belonging to the yellow cluster are mostly related to the Spiccato pieces. The most spread cluster, the light green one, is mostly composed by segments belonging to Martelé recordings. Our hypothesis, that needs to be investigated in the future with an extension of the presented work, is that the set of features we computed on the trajectories can effectively be used to distinguish different bowing exercises and articulation techniques.

6.7.2 Shoulders locked and muscular tension prevention

A second kind of analysis we performed on the data archive of TELMI, was related to muscular tension and rigidity of shoulders movement. We presented this preliminary study in [Volta et al., 2018b], presented at the MOCO conference in 2018. In literature, the interest in music performance has grown, considering several contributions from neuroscience, e.g., [Isabelle Peretz, 2003, Atherton et al., 2018, James and Reifinger, 2018, Gruhn and Rauscher, 2002] to music pedagogy, e.g., [Macrides and Angeli, 2018, Thibeault, 2018, Thompson, 2018] exploring music perfor-

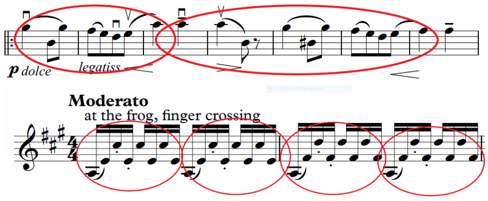



Figure 6.15: Example of the segmentation we made to obtain the evaluation stimuli

mance and learning. An emerging literature investigates how technology, and in particular full-body and motion analysis technologies, may enhance music performance and learning outcomes, by minimizing at the same time the risk of injuries [Marquez-Borbon, 2016], [Peter Visentin and Wasiak, 2008]. In addition, traditional teaching methods of the biomechanics components of musical performance may be based on subjective and vague perception, rather than on accurate understanding of the principles of human movement [Brandfonbrener, 2004]. Musical performance, then, shares many characteristics, including health risks, in common with other skill-oriented activities [Robson, 2004]. It is reasonable to postulate that some methodologies successfully used in sports medicine could be useful in studying the biomechanical aspects of music performance. These understandings include the motor learning theory, learning models, and technology-based system to analyse and validate the effects of training [Magill., 2001]. It is, e.g., a common expectation that students engage in long hours of repetitive practice to acquire technical skills, often not considering the efficacy of such practice and its health implications due to many repetitions of the same movements. Despite many individual success stories, documented rates of injuries among musicians challenge traditional methods [Fishbein et al., 1988]. This study was based on two grounded concepts: motor learning and motor control.

In [Ballreich and Baumann, 1996], theory of motor learning identifies three elements needed to success: the characterization of the skills to be acquired; skills transfer between dissimilar systems, and skills acquisition without injuries. Characterizing skills to be acquired involves a scientific analysis and understanding, to identify and describe motor control patterns, such as the coordination of neural and muscoloskeletal systems. The motor patterns of professional players can be generalized and used as references for a model, to facilitate the identification of skills to be transferred from professional to student musicians. By directing attention to specific motions patterns, learners can assimilate the new skills into their technique efficiently and effectively. In our study, we analysed the data of a multimodal archive of recordings to study movement patterns of shoulders, elbows, and hips. Looking at the variation of angles between such parts of the body, we hypothesized how muscular and posture patterns affect music performance in terms of pitch intonation and dynamic. This study aimed at exploring whether methodologies and technologies from movement analysis can be applied to investigate aspects of the biomechanics of violin performance. These methodologies applied to music performance and learning is not completely new. In [Wöllner et al., 2012] the authors applied motion capture technologies and quantitative analysis of motion to the analysis of prototypical gestures in music conduction.

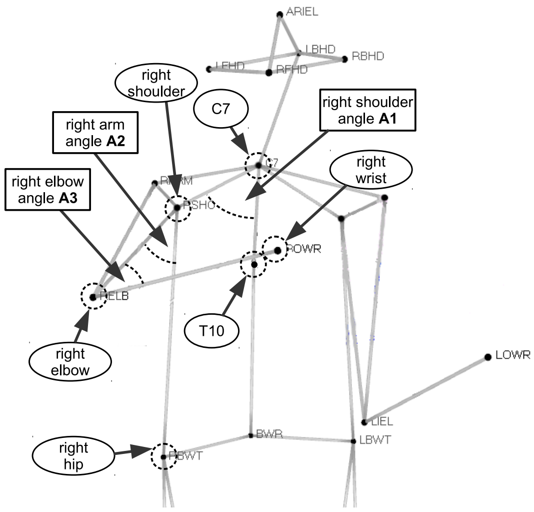


Figure 6.16: The joints (written in rounds) and angles (written in rectangles) we exploit for extracting movement features.

We exploited automated movement analysis to capture the biomechanical skills needed for a physically accurate violin performance, aiming at identifying features able to maximize efficiency and minimize injuries. We focused on 5 players (four of which were the RCM professional violinists recordings) performing three selected exercises (Octave shift from Yost System for Violin, String Crossing from Kreutzer Etude n.13, and Salut d'amour by Edgar, see Figure 6.15). We asked three experts to evaluate the recorded performance for the three exercises, by providing their ratings for 2 audio (i.e., intonation and note production) and 3 movement (i.e., shoulders dynamic and position, and trunk dynamic) features. We finally explored whether and how movement features extracted from full-body motion captured data correlate with the ratings provided by experts. The violinists received the entire list of exercises in advance and the use of music sheets was allowed. The musicians were 4 professional players and one high-level non-professional violinist. After the recordings, the data was segmented, using audio and music sheet information, to isolate single music phrases. In such a way, we extracted 21 segments from the twelve recordings we considered. Using the EyesWeb XMI platform, we computed the measures of the angles between right shoulder and elbow, left shoulder and elbow, the two shoulders, and shoulder-elbow and hip. Considering the theoretical model described in [Peter Visentin and Wasiak, 2008], we focused our analysis on full-body movement features, corresponding to the ones taken into account by violin teachers to evaluate their student's performance.

According to such model, in order to achieve a better performance and minimize the rick of injuries, one could ask a performer "to consciously choreograph movements that minimize static conditions or postures". Moreover, "left elbow should have a degree of lateral movement to

Table 6.1: Spearman's correlations between the ratings provided by the raters and the automatically computed features. Significance level: * for $p < 0.05$, ** for $p < 0.01$

Stimulus	**Shoulders' dynamics**		**Body joints' angles**		**Trunk's dynamics**	
	rho	p-value	rho	p-value	rho	p-value
Audio	.43	.05*	.04	.87	-.03	.91
Video	.55	.008**	-.24	.28	.11	.63
Audio-Video	.61	.005**	-.07	.77	.17	.46

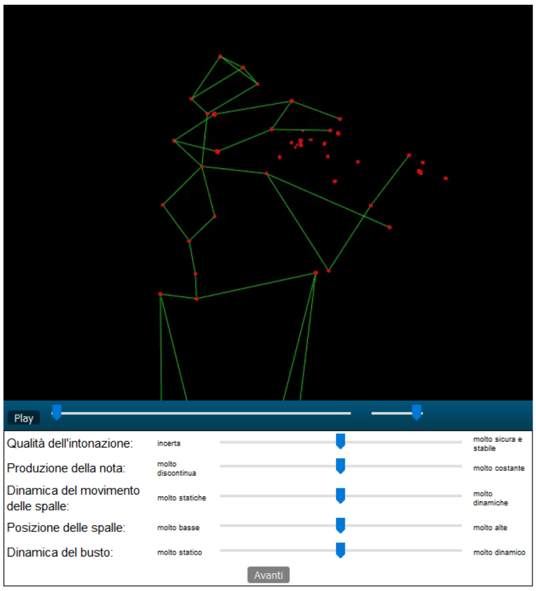


Figure 6.17: Web interface for evaluation of biomechanical violin playing skills: ”Quality of intonation” (”Qualità dell'intonazione”), ”Note production” (”Produzione della nota”), ”Shoulders' dynamics” (”Dinamica del movimento delle spalle”), ”Shoulders' position” (”Posizione delle spalle”), ”Trunk's dynamics” (”Dinamica del busto”)

accommodate playing on different strings”. Also, players that tend to assume unbalanced (and physically risky) postures usually ”locked their shoulders in a highly abducted position”. Some expert violin teachers we interviewed before computing movement features confirmed that “the fight against the tendency to raise one's shoulders is a constant in the teaching, as well as an essential prophylaxis against the onset of very boring and often very damaging professional diseases”. Provided the above background experience, we identified a small initial set of movement features mainly based on (i) angles between joints and (ii) energy of joints movement. The joints and angles we looked at are highlighted in Figure 6.16 and include:

- joints: C7 (base of the neck), T10 (trunk vertebra), shoulder, elbow, wrist;

- angles:
  - the angle between the line C7-T10 and the line C7-shoulder (A1 in Figure 6.16);
  - the angle between the line shoulder-hip and the line shoulder-elbow (A2 in Figure 6.16);
  - the angle between the line shoulder-elbow and the line elbow-wrist (A3 in Figure 6.16).

The above joints and angles are considered both for the right and the left side of the player's body (i.e., angles A1-3 will be computed on both sides). The 3 movement features we compute on the above joints and angles are: **F1, shoulders dynamic** - we compute the Kinetic Energy of left/right shoulder and the first derivative of angles A2 and A3, and we sum up these 3 quantities; the mean between the right and the left side of the body is then calculated, obtaining a single value for F1; **F2, shoulders position** - we compute the right/left mean angle A1; **F3, upper body dynamic** - we sum up the Kinetic Energy of upper body joints C7, T10, and right/left shoulder. F1-3 are computed frame-by-frame on the 21 segments described in the previous section. In this exploratory work, we computed the average of F1-3 on each segment, obtaining a single value of each feature for that segment. Three expert musicians, two violin teachers with more than 20 years of experience in teaching and a professional violin performer, were recruited via email advertisements at academies of music in Italy and France. They were asked to mark their ratings on violin skills, through a questionnaire administered via an ad-hoc web interface (see Figure 6.17). The raters provided their answers to each question by dragging an interactive slider having a resolution of 101 values (the experts did not perceive these values, they perceived a continuous evaluation). The questionnaire consisted of 5 bipolar items ($\alpha = 0.80$) arranged in two sub-scales: a sub-scale related to the acoustic component of violin playing skills (2 items, $\alpha = 0.72$), and a sub-scale related to the movement component of violin playing skills (3 items, $\alpha = 0.70$). Internal consistency was assessed through Cronbach's α. Results show an acceptable reliability both for the whole scale and for each sub-scale [Cronbach, 1951]. The items were: pitch intonation and note production for audio, shoulders' dynamics and position and trunk's dynamics for video. 21 audio/video and audio-video stimuli for a total amount of 63 stimuli were administered in a randomized way. When the stimulus was only audio or video, raters were asked to mark their answers only on the corresponding audio or video sub-scale. Raters were explicitly asked to fully complete the questionnaire. The ratings given by each rater were summed up together over the items.

Inter-rater reliability was assessed through a two-way random, consistency, average-measures ICC [McGraw and Wong, 1996]. It enables to assess the degree at which the 3 raters provided their agreement on violin playing skills across stimuli. The results obtained for the whole scale and each sub-scale were ICC=0.81 (audio-video, 95% CI [0.60,0.92], F(19,38) = 5.24 ,$p < .001$), ICC=0.82 (video, 95% CI [0.54,0,90], F(21,24.4) = 5.68, $p < .001$), ICC= 0.72 (audio, 95% CI [0.24 -0.85], F(20,19.1) = 3.59, $p < .001$), respectively. These high values, falling in the range good and excellent (Cicchetti 1994), indicate that a minimal amount of measurement error was introduced by the 3 independent raters. The confidence interval for audio is very large and one possible explanation we found rely on the cognitive dynamic of audio perception: since the audio sensory modality, in typical people, is informed by visual cues, the absence of visual stimuli could lead to more unsure answers during the evaluation [Boltz et al., 2009]. Pearson's and Spearman's correlations were run to determine the relationship between the perceived violin playing skills and the automatically computed movement features. The analysis showed that there was a strong positive correlation between the shoulders' dynamics and the perceived violin playing skills, independently of the modality of the stimuli. Spearman's correlations showed slightly higher values meaning that there is some deviation from a linear increasing monotonic relationship between the variables. Table 6.1 summarizes the results of the Spearman's correlations for each automated movement features.

Results showed a significant correlation between the evaluation of the experts on the violin playing skills perceived through the different modalities and the measured feature on shoulders' dynamics. This was promising in view of designing systems for music teaching and learning support. Moreover, results showed that all the evaluated modalities (the visual one in particular) can provide useful feedback for both teachers and students on the biomechanical skills needed for a good violin performance. Improvements can be expected, since some results are not significant. For example, the measure of shoulders' position did not correlate with the perceived violin playing skills. This can be motivated by observing that all human annotators highlighted that the violinists exhibited high-locked shoulders. In the future, a similar evaluation could be repeated, using more recordings of different types and violinists to check the extent to which results generalize. Moreover, shoulders position seemed to be performance-dependent, in the sense that the right shoulder rotates vertically according to the string that is currently played. Our algorithm does not take this into account, while human observers did it during the evaluation. The last feature, trunk's dynamics, did not correlate with the experts' ratings too. This result can be explained by the fact that violinists movements during the recordings were limited by the mocap suite and EMG sensors. This non ecological setup could have led to less dynamic movements. In the future, besides improving our algorithms and repeating the evaluation with violinists with more variable skills, we also plan to extend the evaluation, using a more ecological setup, based on portable and non-intrusive devices, such as range imaging sensor.

6.8 SkyMotion: a UI for motion control

Fig. 6.18 shows the overall logical architecture of the TELMI platform. On the top, a remote repository is represented, including an archive of exercises, performances, and performance sessions. Clients are represented around the remote repository, the lower part of the figure representing the fully-featured TELMI client. Blue arrows in the figure represent streams of data the components of the platform exchange. Dashed green arrows represent parameters controlling the processing of single components of the platform. Wide orange arrows represent data from/to the users. Some connections are omitted for sake of clarity. The role of each single component can be briefly summarized as follows:

- Input devices and virtual sensors component: this component is responsible for capturing and pre-processing of multimodal data. This is captured by a broad palette of sensors (e.g. Qualysis motion capture system, Kinect, environmental and pick-up audio, MYO)
- Analysis component: this component is responsible for extracting features from the music performance data. Analysis concerned audio, bow controls, body postures and movements and was organised on multiple layers;
- Real-time feedback component: this component takes as input the values of the features computed above and generates a real-time feedback to be presented to the performer.
- Output devices component: this receives the feedback representations produced by the real-time feedback component and generates the corresponding outputs.
- Recording component: this is responsible for the synchronised recording of all the data collected, computed, or generated during a violin performance with the TELMI platform. The recording component stores data in a local storage device and produces appropriate metadata to identify the performance.
- Offline feedback component: this component is responsible for providing the user (either student or teacher) with views on pre-recorded performance data.
- Session management component: this component is responsible for keeping track of the sessions users perform.
- Teacher interface: this is used by teachers to manage sessions with students.
- Student interface: this is used by students to manage their sessions with the platform.

Concerning the student and teacher interface, the work focused mainly on the development of a new interface for teachers and students for accessing, displaying, and analysing movement data. RCM provided us with the requirements for the interface; we developed it, and integrated it in the final version of the TELMI platform. The interface runs on the top and controls an instance of the EyesWeb XMI platform, endowed with TELMI-specific modules, which performs all the required processing. Fig. 6.19 shows a snapshot of the interface to play back previously recorded movement data and for running specific analyses on such data. In the bottom left corner of the interface, buttons allow the user to choose the data folder and the file to play back. Once a file is selected, playback can be controlled by using the buttons (play, pause, move forward, move backward) on the bottom of the interface, at the centre. A snapshot can be taken to save frames

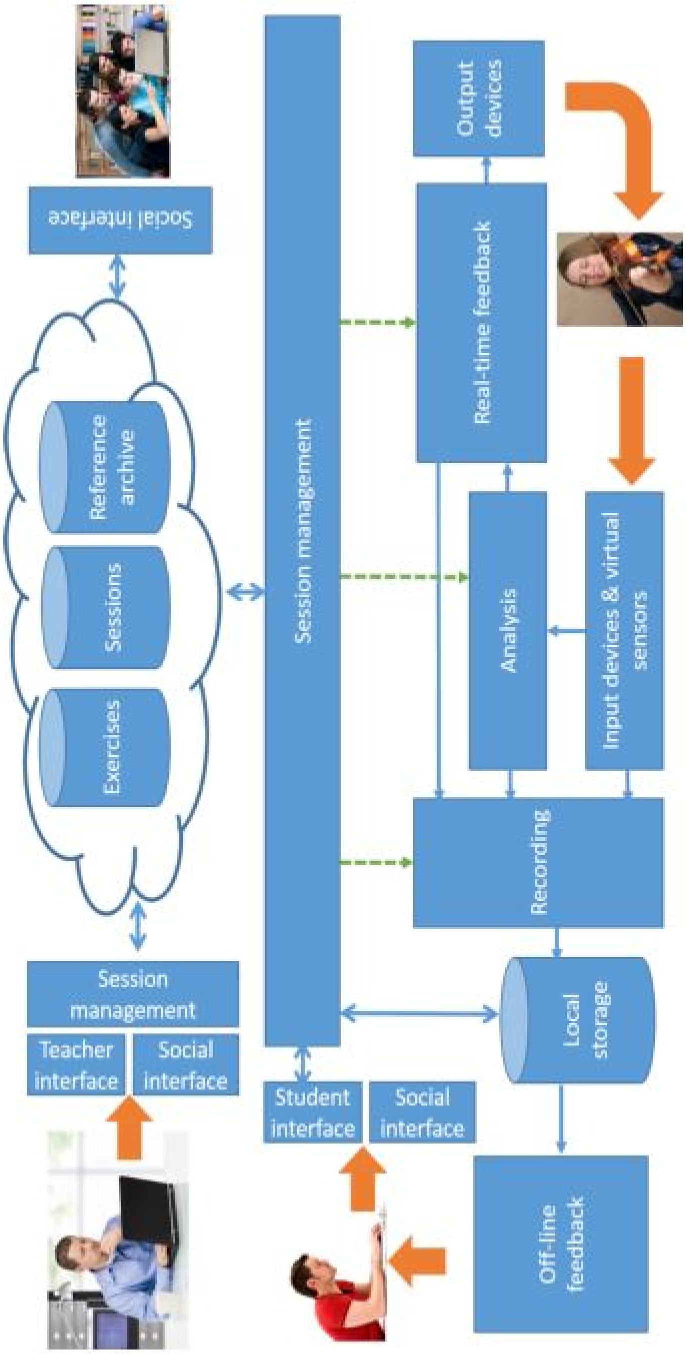
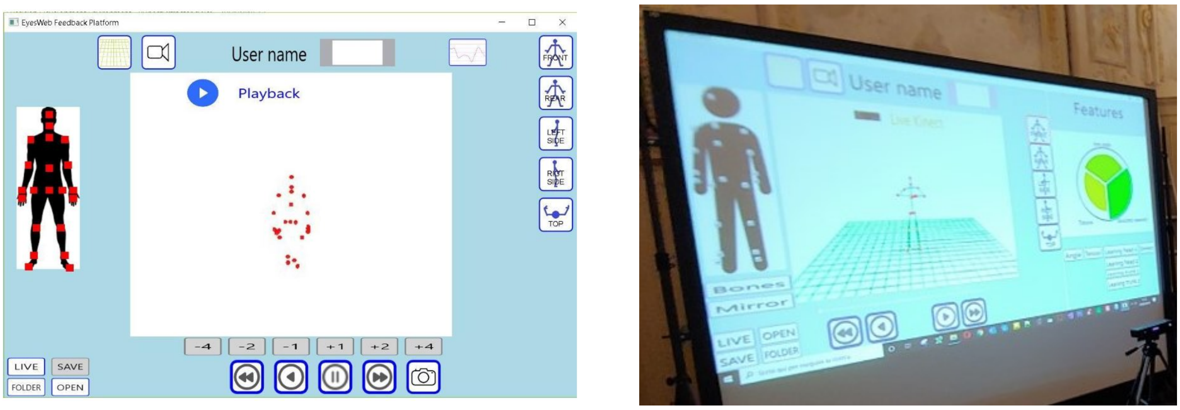


Figure 6.18: Overall logical architecture of the TELMI Platform



(a)

(b)

Figure 6.19: A new teacher and student interface for accessing, displaying, and analysing pre-recorded or real-time movement data. On the left the skeleton with joints that can be selected to be visible or not in the real-time skeleton interface (centre). On the right most suitable features, e.g. tension, bowing and arm movement are monitored with a real-time green to red visualisation to help students focus on bad posture and habits. The interface was developed upon requirements provided by RCM.

that are deemed particularly important (e.g., for further analysis by the teacher). The interface is designed for displaying motion capture data, recorded or displayed in real-time from a Kinect sensor.

Checkboxes located on the human silhouette appearing on the left of the interface allow the user to select the joints to be displayed. By default, all of them are checked and a frontal view of the recorded data is provided in the window at the centre. The viewpoint can however be changed, and the buttons on the right side of the interface enable switching between some pre-set viewpoints: from the front, the back, the left, the right, and the top. Such different viewpoints allow the user (e.g., the teacher) to focus on specific aspects of the violin performance (e.g., the way to keep the violin from the left side, bowing movements from the right side, oscillations and leaning of head and shoulders from the top and so on). Moreover, the interface supports a moving viewpoint, i.e., the viewpoint can move in real-time around the performer enabling the user to look at the performer from different angles (buttons on the top left side of the interface). Finally, movement features can be computed on the displayed movement data: the button on the top right side of the interface, near to the user name, opens dialog boxes for selecting movement features, displaying their graphs in real time, and performing analyses. Fig. 6.20 shows two examples of analysis, separated analysis of upper and lower body on the left, and the graph of the speed of the head on the right. Besides playback of pre-recorded motion capture data, the new interface also supports real-time display and analysis of live data, captured by using a Microsoft Kinect device (or possibly other RGB-D devices supported by the EyesWeb XMI and the TELMI platforms). Fig. 6.19 shows the interface displaying live Kinect data. When running on live Kinect data the interface provides the user with the same functionalities it provides while playing back pre-recorded data. That is, the user can select a subset of joints to focus on and can change such subset in real-time, she can change the viewpoint or move around the performer, she can compute features, perform analyses, and see the graphs of the computed features in real-time.

6.9 Evaluation

In order to provide an evaluation of the prototype developed, a first stage was devoted to understand the parameters on which human assessment is founded. To this end, the following section we outline students' goals for success with relation to the skills necessary for learning the violin and the tools by which these goals will be evaluated. These fall into two categories: human and automatic assessments.

6.9.1 Human assessments

Human assessments, provided by third-party experts and/or as student self-reports, cover the following categories:

- Performance Quality
- Learning Efficiency
- Self-efficacy
- Body use / Exertion

We mainly focused on self-assessments by the performers themselves. This allowed for the study of the degree to which students' reports agree with those of expert assessors, either in overall quality of performance or in improvement over time. The purpose of these assessments was to be to determine the degree to which students using the system (and those not using the system in the experimental control groups) meet performance goals using self-report criteria in line with those used in their day-to-day musical practice. They could be used to give measures of qualities too complex or subjective for automatic assessments (e.g. musicality), opening possibilities for the use of machine learning to determine which automatic components contribute to these factors.

To assess the efficacy of SkyMotion, we created a semi-structured interview (see Appendix B.2 B.2) for seven violin players and teachers that came to Casa Paganini - InfoMus Lab to test SkyMotion and be interviewed in the months between March and September 2019. Through the interviews, the team gathered information to help understanding usability perception and needs

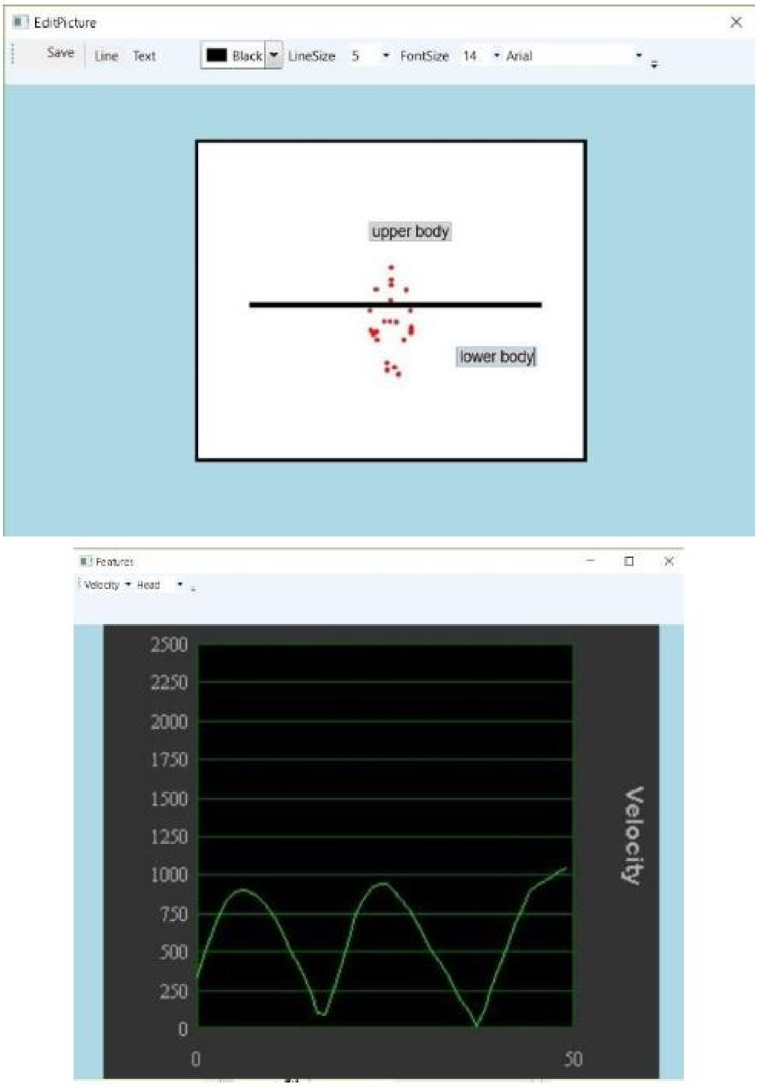


Figure 6.20: An example of analyses the new interface enables: on the left, upper body and lower body are analysed separately (features can extracted separately and compared); on the right the speed of the head is computed and displayed in real-time. Features are computed by the underlying EyesWeb XMI Platform.

understanding of the supposed final group of users, teachers and students of music schools and Conservatoires. The questionnaire was composed of 15 1-7 Likert scale questions, and 4 open questions and was administered to the violinists after they try SkyMotion, playing with no time or repertoire restriction, to independently test the system. The discussion was designed to gather information from users in regard to the following topics:

1. General considerations about the use of learning technologies in class environment and as a support of home practice for students.
2. General considerations about kinesthetic involvement during music practice and performance and what kind of understandings are necessary, from teachers point of view, to progress in the music career.
3. General considerations about what kind of benefit full-body technology can introduce comparing them with technologies and product already on the market.
4. Participants considerations about their perception and understanding of TELMI Full-body interface: points of strengths and weakness, possible problems in mechanic understanding, limits of the actual technology, missing points.
5. Role-play: if the participants were part of a user interface designers team of such a kind of technology what issues they were considered to develop something useful and innovative for their students and what strategies they hypothetically adopt to ensure students engagement and motivation in adopting and using such kind of product.

Seven people took part to the semi-structured interviews: five women and two men. Two were mid-school music teachers and violin performers, one of them already involved in TELMI Project and Suzuki music teacher; one mid-school music teacher and orchestral violin performer, one non-professional violin player, already involved in TELMI Project; two teachers and violin performers, one with orchestral experience, and one non-professional violin performer.

Outcomes from the interviews:

- The TELMI Full-body technology inspires teachers interviewed to explore new methodologies for music learning.
- The TELMI Full-body interface can be used to successfully enhance students' perception on their own body.
- The TELMI Full-body interface let teachers to work on specific postural and movement objectives in a way difficult to do without the technology.

- The TELMI Full-body interface enhance students chance to have real-time feedbacks on their practice also at home, after music classes. These things can improve the development of good skills, in less time than without this home monitoring of the music sessions.
- The TELMI Full-body interface has several strengths: it is a sort of augmented mirror that both students and teachers can use to assess kinesthetic elements difficult to monitor without this technology.
- The novelty of such prospective on music learning is also the element of innovation compared with software already available in market.
- Weaknesses of the TELMI Full-body interface.
- Recommendation for future development.

6.9.1.1 Users Perspectives

6.9.1.2 Answers to open questions from the questionnaires

Question 1: What were the most helpful features?
Expert 1: *"Shoulders symmetry and trunk-elbow angle"*
Expert 2: *"Real-time feedback and the use of colors as augmentation of own proprioception"*
Expert 3: *"Real-time feedback that let the user sees immediately what happened to his/her body"*
Expert 4: *"Real-time feedback on postures"*
Expert 5: *"In general all biomechanical features are helpful, I would focus on very simple but transversal characteristics, e.g. elbow and wrist movements, hand movements, muscular balance between arms and shoulders"*
Expert 6: *"Shoulders monitoring, posture in general"*
Expert 7: *"Positive aspect is that it force me to think about posture and focus on that while playing"*

Question 2: What were the biggest weaknesses/limitations?
Expert 1: *"Limited numbers of features"*
Expert 2: *"The swift of attention required to pay attention to the interface while I was playing"*
Expert 3: *"Not so easy to use, especially considering young learners"*
Expert 4: *"Not so precise in detecting small body parts position, e.g. wrist or hand"*
Expert 5: *"It not enough precise on small movements, fundamental for violinists, e.g. wrists movements and hands articulation"*
Expert 6: –
Expert 7: "Not enough accurate in body joints detection and measurement"

Question 3: How would you improve the feedback? Expert 1: *"I would like to insert an audio feedback on movements and posture analysis"*
Expert 2: *"Probably I would like to design the interface to be less reactive, taking more time to color all the shape. In this way the user would have more time to pay attention to the interface and then to his/her own position."*
Expert 3: *"I would like to add back view of the body to be able to monitor posterior muscular chains"*
Expert 4: *"I would like to have a focus on hand movements and positions"* Expert 5: *"I would like to add some other measurements that let to monitor wrist and hands movements"* Expert 6: –
Expert 7: *"I would like to add more parameters on shoulder posture, not just stiffness of position to consider its tension"*

Question 4: Any other feedback on the technology?
Expert 6: *"Useful for technics control"*
Expert 7: *"Reduce precision problems due to particular combination of dress color that may interface with the Kinect system"*

Outcome 1: The Full-body technology inspires teachers interviewed to explore new methodologies for music learning. **Question asked during interviews:** "Does the project inspire you to try out new methodologies in your violin teaching?" **Findings from the interviews:** all the teachers interviewed expressed interest in trying and include in their own practice new methodologies of teaching violin learning. The use of TELMI technology inspired them to imagine using such a kind of technology and how to include it in their teaching routing.

> *"Posture is very important, but, especially when students are young and start to study, I have encountered major problems with posture, because many hours of study can lead to bad postural habits and pain difficult to eliminate. Such a kind of methodology, that can focus of naturalness of the violinist body which she is playing can be really interesting to be included in violin pedagogy."*

> *"One of the aspect that would let me to deepen new methodologies of teaching, using SkyMotion, is the form of gamification that using this kind of technology is possible to introduce in the lesson."*

> *"SkyMotion is an improvement compared with traditional self-learning methods, as study in front a mirror is, it is a kind of augmented feedback that let students to focus their attention on tensions that not always are easy to see in front of the mirror"*

Outcome 2: The TELMI Full-body interface can be used to successfully enhance students' perception on their own body. **Question asked during interviews:** "What in your opinion is important competence, from a biomechanical point of view, in music performance?" **Findings from the interviews:** Teaching to be aware of body during the musical activity, to reach and maintain the naturalness of posture and body during the performance.

> *"Students usually want to grab the bow, hold the violin with force, stretching the cervical, raise and stiffen the shoulder, because they are afraid that the violin falls. In the same way, I insist a lot on the legs position because some students cross the legs while they play, in this way the base is missing. If, in the face of all this, I insist that the student concentrate on a natural position, let fall the natural weight of the arms, then raise the arm, I notice that the students take fewer defects. So keep the position as natural as possible, thinking, especially if they are so young, to the right position of the finger, which in turn links the other correct postures: that of the wrist, elbow, etc.. For the bow, the speech is the same: if a person lets go of her/his arm naturally, the thumb itself goes into the correct position. So I think, the competence to teach is an awareness of the naturalness and softness of movement therefore."*

> *"The most important aspect from a mechanical point of view is to coordinate own breath with movements while playing, because in this way it is possible to play without muscular tension. Unfortunately, is one of the most difficult violin aspect to master, also after several years of study. SkyMotion would help students to understand the importance of posture for a good playing instead of completely focused just on sound production."*

Outcome 3: The TELMI Full-body interface let teachers to work on specific postural and movement objectives in a way difficult to do without the technology. **Question asked during interviews:** "What would be your goals as a violin teacher, if you consider using such a kind of technology?" **Findings from the interviews:** Teaching to be aware of body during the musical activity, to reach and maintain the naturalness of posture and body during the performance.

> *"Learning to limit body tension, because when a performance lasts for long period, getting to the end with stiffness is a huge problem and it's of fundamental importance to teach good habits to students."*

> *"I would like to use SkyMotion with my students to help them understand the importance of working on own body before working on music performance."*

Outcome 4: The TELMI Full-body interface enhance students chance to have real-time feedbacks on their practice also at home, after music classes. These things can improve the development of good skills, in less time than without this home monitoring of the music sessions.

Question asked during interviews: “We’d like to hear about the classes in which you think this technology may be successfully applied. In what ways may the interface be helpful to you and/or in the classes you are considering?” **Findings from the interviews:** The TELMI full-body interface can be very useful for the home student’s self-monitoring, during the week of study, so that s/he can realize and correct bad postures or movements, stiffening and bad habits, before returning to music class and then progress in a way that is faster and safer for her/him in the learning process. In class, TELMI interface can be used, but rarely, more to show something specific to the student. This is also due to the fact that lesson time is short and therefore it would not be useful to use it continuously and for the whole time.

> *“It can be very useful for students at home, since after hours of practice they are tired and easily lose their correct posture and its naturalness, increasing muscular tensions.”*

> *“As it is a kind of augmented mirror, it would be really useful for students for self-learning and monitoring at home while studying. In particular, I would focus just of mechanical aspects of violin playing, such as articulation of left hand, wrists positions and movements (left wrist shifting for example, very important for a good vibrato starting from the wrist). However, to be really helpful, it should be extremely precise, even on very small variation of movements or posture, otherwise it would not help so much.”*

Outcome 5: The TELMI Full-body interface has several strengths: it is a sort of augmented mirror that both students and teachers can use to assess kinesthetic elements difficult to monitor without this technology. The novelty of such prospective on music learning is also the element of innovation compared with software already available in market. **Question asked during interviews:** “What are the strength points of the system?” **Findings from the interviews:** The TELMI full-body interface is a sort of augmented mirror, let students focus their attention on movements generally unconscious or difficult to monitor. It is an important tool to teach body posture and its naturalness during music playing. It can help also students understanding the relation between body posture and correct movements and clearness and precision of sound production and on the listening itself. It can be very helpful on techniques studies, even if it is not suitable for students under 11-12 years old. It can help to eliminate physical tiredness, since one of the major problem of the students group considered (from 11-12 years old) is to study several hours per day, so diminishing tension (especially for shoulders) can lead to a better sound production. Furthermore, for a proper use of the interface, students should learn by heart the music piece and this will be another fundamental, even if indirect, positive effect of its use. Student awareness and being an extra monitoring tool for teacher are the two main strengths of the system.

"Music is not in the written page, but in the musical thought, so even if it is a side effect, the fact of learning by heart the music to play, to monitor the performance with this technology, would be an effect on a very important educational goal. Moreover, the other objective would be to maintain even in the most complex passages the balance between the movement features that the system monitors, allowing the student to focus on that sense of naturalness of posture that is often to be rediscovered."

Outcome 6: Weakness of the TELMI Full-body interface. **Question asked during interviews:** "What are the weakness of the system?" **Findings from the interviews:** The TELMI full-body interface is not sufficiently applied to violin repertoire. The system representation of the body is not immediately clear and should be improved to be easier to understand also for younger students. There is no, at the actual stage, a way to monitor the bowing techniques and it should be added.

"The disadvantage could be exactly the same as the advantage: if the student focuses too much on monitoring his or her results through the interface, s/he might create tension is her/his body. However, it is an initial disadvantage, given to the fact that s/he has to learn to use the technology, because, as in all other aspects, if I focus too much on one element, in an exasperated way, I could create tensions in the performance."

6.9.1.3 Recommendation for future development

One feature that we have to add is tension (both muscular and postural) of lower part of the body, in particular of hips and legs. Visual representation of the skeleton should be improved to be more suitable to younger and non-expert students. For one of the violinist would be better to add a 3D avatar representation instead of a skeleton, to better connect the visual feedback with own proprioception. Furthermore, it would be interesting to have the chance to color directly the body joints the system reveal to stretched, to further enhance the comprehension of own posture. Moreover, it could be really interesting to have a system that adapts feedback based also on the type of repertoire and particular needs that the repertoire highlights. It can be also really useful to have a records archive both for students and teacher monitoring, so that the students can have a time perception of their improvements while the teacher can monitor home study results and maybe adapt her/his learning or adding suggestion specifically based on body monitoring recordings. Considering this point, also a representation of the performance trend over time should be added and also a bowing techniques monitoring should be considered, since it is a particularly important element of violin pedagogy. It may be interesting to consider profiling, from the point of view of the movement patterns, of great musicians, perhaps on specific repertoires, to create a sort of pattern of reference for the student who is facing the same repertoire. The system should also include some audio features to let student better integrate body awareness with sound production. One core aspect would be to have the chance to better monitor hands and wrists position and movement, to let SKyMotion be truly useful for students. Finally, one thing that all the teachers highlight is that could be nice to have also the possibility to add own music sheet to adapt the system in time to actual repertoire needs.

6.9.1.4 Suggestions for future research

A research study in which using this kind of technology, it can be possible to measure the quality of the performance, both from a postural and sound point of view, comparing different kind of techniques, already existing, to train musicians' body to enhance their awareness and improve their sound results.

6.9.1.5 Results from the questionnaire

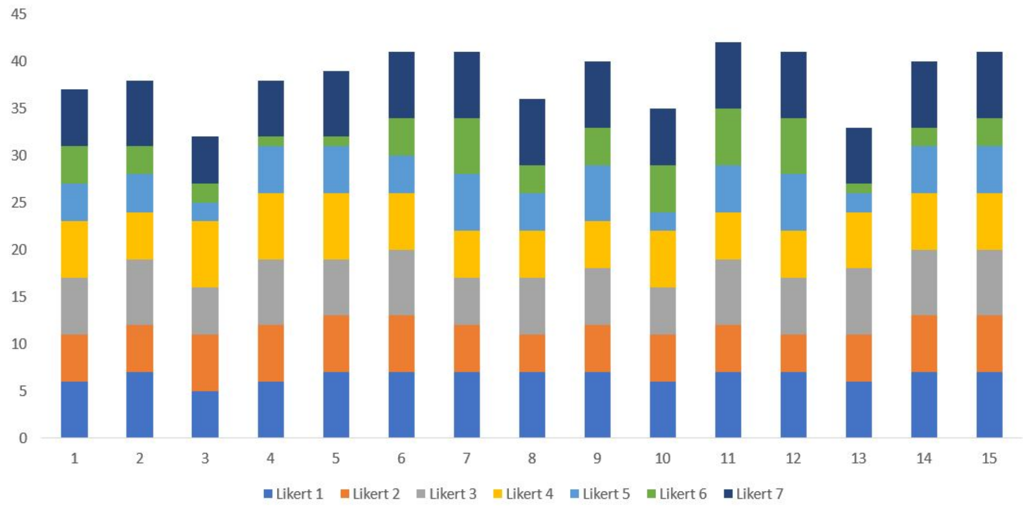
Here we presented the results of the Likert questionnaires given to the violinists (see Table 6.2. In Fig. 6.21 we presented the results from the Likert questionnaires given by frequency of annotations. As it can be seen, annotation ratings were very similar between the violinists.

Table 6.2

No.	Question	Expert Likert evaluation						
	To what degree did this feedback…	**Exp.1**	**Exp.2**	**Exp.3**	**Exp.4**	**Exp.5**	**Exp.6**	**Exp.7**
1	…help you learn more quickly?	6	5	6	6	4	4	6
2	… improve your performance?	7	5	7	5	4	3	7
3	… increase your productivity?	5	6	5	7	2	2	5
4	… increase the effectiveness of your practice?	6	6	7	7	5	1	6
5	… make practicing easier?	7	6	6	7	5	1	7
	To what degree was this feedback…							
6	… useful?	7	6	7	6	4	4	7
7	… easy to learn to operate?	7	5	5	5	6	6	7
8	… something that did what you wanted it to do?	7	4	6	5	4	3	7
9	… clear and understandable?	7	5	6	5	6	4	7
10	… flexible?	6	5	5	6	2	5	6
11	… easy to become skilled at?	7	5	7	5	5	6	7
12	… easy to use?	7	4	6	5	6	6	7
13	… accurate?	6	5	7	6	2	1	6
14	… something you would use again?	7	6	7	6	5	2	7
15	… something you would recommend to others?	7	6	7	6	5	3	7

To conclude, here we presented a semi-structured questionnaire for interviewing violin performers and teachers on their experience using the SkyMotion interface. Results are quite promising, since all the violinists interviewed agreed on finding useful a full-body interface that may help students understanding the difficult biomechanics behind violin learning. Of course, they also

Figure 6.21: The figure shows the frequency of annotation of Likert values by the seven violin players interviewed of the 15-items Likert questionnaire presented in Appendix B.2



highlighted further improvements the system should consider, as a deepen monitoring of very small and precise articulation and postures of specific joints, e.g. wrist and hands.

Chapter 7

Conclusion and future steps

The literature examined the role of multisensoriality and embodiment in learning, considering the most recent research in neuroscience, psychology, pedagogy and how they can direct HCI towards new definitions of multimodal systems. In particular in this work we have focused on two different kinds of learning: (a) learning in the age of development, both in the presence or absence of sensory deficits such as visual and learning deficits (dyslexia) and (b) a second type of learning, the instrumental music learning, namely a multimodal by definition and transversal in terms of age group. In the developmental age, the main focus was on learning mathematics in primary school, exploiting new solutions for children's interaction with technology, understanding the role of proprioception and the sound-musical channel to create a play experience that intuitively refers to mathematical properties took into account. Central and extremely important element in the design developed is the concept of inclusive technology, so that the same technological solution should be usable independently by both children with typical development and those with sensory deficits. The main objective of this work was therefore to investigate, through empirical applications to concrete case studies, the design and development of multisensory technologies, based on a more careful examination of the functioning of sensory modes in the different groups of users considered. In particular, the first case study, based on the European Project H2020 weDRAW, has allowed to deepen - thanks to the team work with the partners of University College London, Knlowedge Lab and UCLIC, and the partners of UVIP team from the Italian Institute of Technology -, the difficulties in learning the mathematics of English and Italian primary school children, both in typical development and with sensory deficits, and to understand which are the preferential sensory channels in learning specific concepts, so as to develop prototypes that would respect these findings. In particular, for the weDRAW projects a considerable amount of job was devoted to implement and test multisensory serious-games suitable to teach mathematical concepts to visually impaired children, investigating what role sounds feedback should have to proper reach the goal. Equally important was the development of assessment tools, both at the psychophysical and technology assessment level, that would allow us to understand to what extent the models developed and the prototyped technologies would allow us to achieve our educational goals. Finally, considering specific learning deficits such as dyslexia, it was important to have had the opportunity to develop a multisensory test model that could serve to highlight sensory difficulties at ages far before those in which it is actually possible to recognize and intervene specifically on dyslexia. The work on dyslexia has only been started, but it turned out to be an interesting case study on how to apply multisensory perceptual learning model on learning disabilities. As for the second case study, the H2020 TELMI project, the project aimed at developing learning technologies to support the self-learning of the violin student in his/her path. The pedagogy taken into consideration is the traditional classical pedagogy, which we were able to deepen thanks to the work of the partner Royal College of London. Within the project, studying biomechanics at the basis of the correct body awareness of the musician, allowed us to develop an interface, SkyMotion, in which you can monitor your body during the study. In particular, SkyMotion allows you to monitor specific qualities of movement and posture, such as shoulders symmetry or body tension. As for the weDRAW technologies, we deeply believe in the importance of human evaluation of the technology developed. For this reason, we create a semi-structured questionnaire with which we interviewed seven violin performers and teachers, talking with them about the strengths and weakness of the system we developed and on the chance of developing new pedagogical framework, based on this kind of outcome. Results of these interviews showed SkyMotion as a good prototype and starting point for new learning technology aimed at helping self-learning in complex case, such as violin learning. Finally, in both weDRAW and TELMI case studies, our results highlighted also weakness points that would need more research to deepen them. For example, it would be necessary to further consider non-verbal behaviors both in typically developed children and visually impaired ones, to be able to build a comprehensive library of such non-verbal communication that could let the weDRAW platform to become truly adaptive. Furthermore, the weDRAW games are at a prototype stage and should be further developed, creating multi-levels game narrative, allowing the game be suitable for all the primary school curriculum and not only the first years of study. Considering the second case study and the work done for TELMI, we should use the data collected in the TELMI multimodal archive, to create a model of performance, e.g. "student vs. teacher" model, to be able to monitor with more precision the goodness of the students' performance. Furthermore, we should consider other non-invasive sensors to be integrated into SkyMotion, to be able to monitor small movements of hands and wrists as the violin players interviewed asked, to let the system be useful for students of several grade of study. Moreover, it would also be interesting to deepen our investigation on bad postural habits and posture tension that may lead, if chronicle, to injuries and chronic pain. For that study it would be nice to add physiological data recorded, as EMG Myo data to the analysis. In conclusion, this research has deepened and examined important aspects highlighted by the literature on multisensory perceptual learning, confirmed some results and hypotheses and obtained further considerations for a good technology design that surely deserve to be further investigated in future research.

List of Figures

Appendix A

A.1 Teacher Questionnaire (Italy)

Arithmetic and geometric problems according to age group (elementary)

Age __________________

Teacher (subject) ________________________________

Classes taught previously 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] Media [ ]

Classes I teach now

1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] Media [ ]

Arithmetic

1. Indicate the areas of arithmetic that children have difficulty learning, depending on their class, selecting from the list of topics attached:

	Difficulty 1	Difficulty 2	Difficulty 3	Difficulty 4
Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

2. Describe how the four concepts of difficulty you mentioned above are currently taught:

Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

3. Which sensory modality (vision, hearing, touch) is generally used to teach the four difficult concepts identified?

Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

Geometry

1. Indicate the areas of geometry that children have difficulty learning, depending on their class, selecting from the list of topics attached:

	Difficulty 1	Difficulty 2	Difficulty 3	Difficulty 4
Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

2. Describe how the four concepts of difficulty you mentioned above are currently taught:

Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

3. Which sensory modality (vision, hearing, touch) is generally used to teach the four difficult concepts identified?

Prima elementare (1st year)				
Seconda elementare (2nd year)				
Terza elementare (trd year)				
Quarta elementare (4th year)				
Quinta elementare (5th year)				

GENERAL

1. Most of the arithmetic and geometric concepts are conveyed using the visual mode. However, some forms of disability reduce access to particular sensory channels, as in the case of blindness that cannot exploit the visual channel. To what extent could other sensory modalities be useful for the teaching of the following concepts? Enter a percentage value that indicates the contribution of specific sensory modalities in teaching specific concepts.

	Visual	Audio	Touch	Tot.
Number lines				
Symmetry				
Shapes and spatial relationships				
Measurement and estimation				

2. Indicate the level of difficulty in learning the concepts listed below, depending on the class and consequently, the age group.

	Difficulties									
	1	2	3	4	5	6	7	8	9	10
Prima elementare (1st year)										
Number lines										
Symmetry										
Shapes and spatial relationships										
Measurement and estimation										
Seconda elementare (2nd year)										
Number lines										
Symmetry										
Shapes and spatial relationships										
Measurement and estimation										
Terza elementare (trd year)										
Number lines										
Symmetry										
Shapes and spatial relationships										
Measurement and estimation										
Quarta elementare (4th year)										
Number lines										
Symmetry										
Shapes and spatial relationships										
Measurement and estimation										
Quinta elementare (5th year)										
Number lines										
Symmetry										
Shapes and spatial relationships										
Measurement and estimation										

List of possible topics to be used for the questionnaire: You can use topics not on the list.

Arithmetic

- Knowing numbers and the ways of representing them
- Establishing and representing relationships
- Train sets and establish relationships (?)
- Predictions
- The majority relations, minority and equality (?)
- Compare and sort numbers
- Addition and subtraction
- The groupings for ten (decimalisation?)
- Classification of elements (?)
- The groupings (?)
- Ordinances in general
- Ordinances in progressive and regressive sense
- Compare general
- Comparison between quantities and numbers.
- Compare sizes
- Mental Calculation
- Equivalencies
- The meter, the different sizes
- Very large natural numbers (composition, decomposition, comparison and sorting of numbers),
- Percentages

Geometry

- Know the earliest forms of plane and space
- Use graphics
- Work on the Cartesian plane
- Isometric transformations
- The spatial indicators
- Orientation: paths and locations (coordinates)
- Identification in the reality of solid geometric shapes.
- Lines and lines (?)
- Straight and their position in space
- Angles: construction, representation and classification
- The meter, the different sizes
- Isometric Transformations: tipping, translation and rotation
- Travel in the Cartesian plane
- Compare sizes

A.2 National Curriculum for Primary Mathematics by Year

NATIONAL CURRICULUM FOR PRIMARY MATHEMATICS BY YEAR

	Fractions	***Measurement***	***Geometry (shapes)***	***Geometry (position)***
Year 1 (age 5)	Recognise, find and name a half as one of two equal parts of an object, shape or quantity. Half and quarter as 'fractions of' discrete and continuous quantities by solving problems using shapes, objects and quantities. For example, recognise and find half a length, quantity, set of objects or shape.	Compare, describe, measure and record, solve practical problems for: length/height; weight/mass; capacity/volume; time. Begin to be familiar with measuring instruments		
Year 2 (age 6)	Write simple fractions for example, 1/2 of 6 = 3 and recognise the equivalence of 2/4 and 1/2. Count in fractions up to 10, starting from any number and using the 1/2 and 2/4 equivalence on the number line (for example, 1 1/4 , 1 2/4 (or 1 1/2), 1 3/4 , 2). Reinforces the concept of fractions as numbers and that they can add up to more than one.	Use appropriate units for estimating and measurement e.g length/height in any direction (m/cm); mass (kg/g); temperature (°C); capacity (litres/ml) Compare and order lengths, mass, volume etc. (comparing measures including simple multiples such as 'half as high'; 'twice as wide').	Identify and describe the properties of 2-D shapes, including the number of sides and line symmetry in a vertical line. Identify and describe the properties of 3-D shapes, including the number of edges, vertices and faces. Identify 2-D shapes on the surface of 3-D shapes, (for example, a circle on a cylinder and a triangle on a pyramid). Compare and sort common 2-D and 3-D shapes and everyday objects.	Use mathematical vocabulary to describe position, direction and movement, including movement in a straight line and distinguishing between rotation as a turn and in terms of right angles for quarter, half and three-quarter turns (clockwise and anti-clockwise). Work with patterns made up of shapes, including those in different orientations. Use the concept and language of angles to describe 'turn' by applying rotations, including in practical contexts (for example, pupils themselves moving in turns, giving instructions to other pupils to do so, and programming robots using instructions given in right angles.
Year 3 (age 7)	Count up and down in tenths; recognise that tenths arise from dividing an object into 10 equal parts and in dividing one digit numbers or quantities by 10.	Measure, compare, add and subtract lengths, mass, Volume. Measure perimeter of 2D shapes. Compare measures including simple		

	Add and subtract fractions with the same denominator within one whole [for example, 5/7 + 1/7 = 6/7]. Compare and order unit fractions, and fractions with the same denominators.	scaling by integers (for example, a given quantity or measure is twice as long or five times as high). This connects to multiplication.		
Year 4 (age 8)	Understand decimals and fractions are different ways of expressing numbers and proportions. Understanding is extended to tenths and then hundredths. This includes relating the decimal notation to division of whole number by 10 and later 100. Practise counting using simple fractions and decimals, both forwards and backwards. Learn decimal notation and the language associated with it, including in the context of measurements. Make comparisons and order decimal amounts and quantities that are expressed to the same number of decimal places. Represent numbers with one or two decimal places in several ways, such as on number lines.	Convert between different units of measure [for example, kilometre to metre; hour to minute] Measure and calculate the perimeter of a rectilinear figure (including squares) in centimetres and metres. Find the area of rectilinear shapes by counting squares Build on understanding of place value and decimal notation to record metric measures, including money. Use multiplication to convert from larger to smaller units. Express perimeter algebraically as 2(a + b) where a and b are the dimensions in the same unit. Relate area to arrays and multiplication.	Extend classification to different triangles (for example, isosceles, equilateral, scalene) and quadrilaterals (for example, parallelogram, rhombus, trapezium). Identify acute and obtuse angles and compare and order angles up to two right angles by size. Decide if a polygon is regular or irregular. Identify lines of symmetry in 2-D shapes, presented in different orientations. Draw symmetric patterns to become familiar with different orientations of lines of symmetry. Recognise line of symmetry in a variety of diagrams, including where the line of symmetry does not dissect the original shape.	Describe positions on a 2D grid as coordinates in the first quadrant. Describe movements between positions as translations of a given unit to the left/right and up/down. Plot specified points and draw sides to complete a given polygon. Draw a pair of axes in one quadrant, with equal scales and integer labels. Read, write and use pairs of coordinates, for example (2, 5), including using coordinate- plotting ICT tools.
Year 5 (age 9)	Compare and order fractions whose denominators are all multiples of the same number. Identify, name and write equivalent fractions of a given fraction, represented visually, including tenths and hundredths. Recognise mixed numbers (e.g. 1 ½) and	Use knowledge of place value and multiplication and division to convert between standard units. Calculate the perimeter of rectangles and related composite shapes, including using the relations of perimeter or area to find unknown lengths. Missing measures questions such as these	Identify 3-D shapes, including cubes and other cuboids, from 2-D representations. Know angles are measured in degrees: estimate and compare acute, obtuse and reflex angles. Draw given angles, and measure them in degrees. Identify: angles at a point and one whole turn (total 360	Identify, describe and represent the position of a shape following a reflection or translation, using the appropriate language, and know that the shape has not changed. Recognise and use reflection and translation in a variety of diagrams, including continuing to use a 2-D grid and coordinates in the first quadrant.

	improper fractions (3/2) and convert from one form to the other. Write mathematical statements > 1 as a mixed number (e.g. 1 ½). Add and subtract fractions with the same denominator, and denominators that are multiples of the same number. Multiply proper fractions and mixed numbers by whole numbers. Read and write decimal numbers as fractions. Recognise and use thousandths and relate them to tenths, hundredths and decimal equivalents. Round decimals with two decimal places to the nearest whole number and to one decimal place. Read, write, order and compare numbers with up to three decimal places. Solve problems involving number up to three decimal places. Recognise the percent symbol (%) and understand that percent relates to 'number of parts per hundred'. Write percentages as a fraction with denominator 100, and as a decimal. Solve problems which require knowing percentage and decimal equivalents of ½ ¼ etc. and those fractions with a denominator of a multiple of 10 or 25.	can be expressed algebraically, for example 4 + 2*b* = 20 for a rectangle of sides 2 cm and *b* cm and perimeter of 20cm. Calculate the area from scale drawings using given measurements.	degree); angles at a point on a straight line and 1/2 a turn (total 180 degrees) . Use the properties of rectangles to deduce related facts and find missing lengths and angles. Distinguish between regular and irregular polygons based on reasoning about equal sides and angles.	
Year 6 (age 10)	Use common factors to simplify fractions; use common multiples to express fractions in the same denomination.	Solve problems involving the calculation and conversion of units of measure, using	Draw 2-D shapes using given dimensions and angles.	

	Compare and order fractions, including fractions > 1. Add and subtract fractions with different denominators and mixed numbers, using the concept of equivalent fractions. Multiply simple pairs of proper fractions, writing the answer in its simplest Form. Divide proper fractions by whole numbers (for example, 1/3 ÷ 2 = 1/6). Associate a fraction with division and calculate decimal fraction equivalents (for example, 0.375) for a simple fraction (for example, 3/8). Identify the value of each digit in numbers given to three decimal places and multiply and divide numbers by 10, 100 and 1000 giving answers up to three decimal places. Multiply one-digit numbers with up to two decimal places by whole numbers. Use written division methods in cases where the answer has up to two decimal places. Solve problems which require answers to be rounded to specified degrees of accuracy. Recall and use equivalences between simple fractions, decimals and percentages, including in different contexts.	decimal notation up to three decimal places where appropriate. Use, read, write and convert between standard units, converting measurements of length, mass, volume and time from a smaller unit of measure to a larger unit, and vice versa, using decimal notation to up to three decimal places. Convert between miles and kilometres. Recognise that shapes with the same areas can have different perimeters and vice versa. Recognise when it is possible to use formulae for area and volume of shapes. Calculate the area of parallelograms and triangles. Calculate, estimate and compare volume of cubes and cuboids using standard units, including cubic centimetres (cm^3) and cubic metres (m^3), and extending to other units (for example, mm^3 and km^3).	Recognise, describe and build simple 3-D shapes, including making nets. Compare and classify geometric shapes based on their properties and sizes and find unknown angles in any triangles, quadrilaterals, and regular polygons. Illustrate and name parts of circles, including radius, diameter and circumference and know that the diameter is twice the radius. Recognise angles where they meet at a point, are on a straight line, or are vertically opposite, and find missing angles. Describe the properties of shapes and explain how unknown angles and lengths can be derived from known measurements. These relationships might be expressed algebraically, for example, $d = 2 \times r$; $a = 180 – (b + c)$.	

Appendix B

B.1 weDRAW Usability Checklist

Usability Evaluation Report

Date

OBSERVER:

GAME UNDER EVALUATION:

USER IDENTIFICATION: **AGE:** **F** **M**

VISUAL IMPAIREMENT: **VISUAL RESIDUE (IF APPLICABLE):**

OTHER RELEVANT INFORMATION:

Check if the player has difficulties to:			*Did it happen?*		*In which task?*	*Severity*
Gameplay	1.1	Learn to play	Y	N		1 2 3 4
	1.2	Learn how to use the controls 1.16 Link what she did with the body with the request of the game	Y	N		1 2 3 4
	1.3	Handle the controls (if applicable)	Y	N		1 2 3 4
	1.4	Understand the game goals (if applicable)	Y	N		1 2 3 4
	1.5	Play without mediation	Y	N		1 2 3 4
	1.6	Play according to the information provided by the game	Y	N		1 2 3 4
	1.7	Accomplish the game tasks	Y	N		1 2 3 4
	1.8	Move through the virtual game environment (if applicable)	Y	N		1 2 3 4
	1.9	Rotate in the virtual game environment (if applicable)	Y	N		1 2 3 4
	1.10	Recognize different scenarios in the game (if applicable)	Y	N		1 2 3 4
	1.11	Distinguish the different characters in the game (if applicable)	Y	N		1 2 3 4
	1.12	Distinguish the distinct roles in the game (if applicable)	Y	N		1 2 3 4
	1.13	Enjoy the interaction	Y	N		1 2 3 4
	1.14	Understand the link with the mathematical concepts of the game	Y	N		1 2 3 4
	1.15	Use the body in the activity	Y	N		1 2 3 4

	1.16	Link what she did with the body with the request of the game	Y	N		1 2 3 4
	Check if the player demonstrates to feel :					
	1.17	Bored, or uninterested while playing	Y	N		1 2 3 4
	1.18	Annoyed by any of the game controls	Y	N		1 2 3 4
	1.19	Confused by the mechanic of the interaction and/or the task of the game	Y	N		1 2 3 4
Acoustics	2.1	Hear the game sounds	Y	N		1 2 3 4
	2.2	Identify a specific sound	Y	N		1 2 3 4
	2.3	Recognize a specific sound	Y	N		1 2 3 4
	2.4	Understand the information conveyed by a sound	Y	N		1 2 3 4
	2.5	Realize that a specific sound is related to a specific action in the game	Y	N		1 2 3 4
	2.6	Associate the game sounds with his prior knowledge	Y	N		1 2 3 4
	2.7	Associate the game sounds with the right objects or actions in the game	Y	N		1 2 3 4
	2.8	Understand information about orientation and location	Y	N		1 2 3 4
	2.9	Identify the purpose of a specific audio feedback	Y	N		1 2 3 4
	2.10	Identify the link between the sounds heard and the movement/action performed	Y	N		1 2 3 4
	2.11	Identify the function of the sounds in the game (e.g.	Y	N		1 2 3 4

		count, a reference for the movement/rotation, a sign of dimension of shapes/angles, …)			
	Check if the audible feedback :				
	2.12	Is sufficient to the execution of the game activities	Y N		1 2 3 4
	2.13	Is correctly applied to the game objects and execution of game activities	Y N		1 2 3 4
	2.14	Is clearly enough to be properly distinguished?	Y N		1 2 3 4
	Check if the player demonstrates to feel :				
	2.15	Uncomfortable with the speed of spoken audio or TTS (Text-To-Speech) (if applicable)	Y N		1 2 3 4
	2.16	Annoyed or scared by specific sounds or voices	Y N		1 2 3 4
ONLY FOR TYPICALLY DEVELOPING CHILDREN AND CHILDREN WITH VISUAL RESIDUES:					
Graphics	3.1	Understand the information conveyed by images	Y N		1 2 3 4
	3.2	Understand the information conveyed by colors	Y N		1 2 3 4
	3.3	See fonts, figures, diagrams or other graphic elements	Y N		1 2 3 4

B.2 Telmi Usability Questionnaire

FOCUS GROUP QUESTIONS – TELMI FULL-BODY INTERFACE

T E L M I
Technology Enhanced Learning of Musical Instrument Performance.

INDIVIDUAL TECHNOLOGY ASSESSMENT

To what degree did this feedback…	**(from 1=low to 7=high)**						
…help you learn more quickly?	1	2	3	4	5	6	7
… improve your performance?	1	2	3	4	5	6	7
… increase your productivity?	1	2	3	4	5	6	7
… increase the effectiveness of your practice?	1	2	3	4	5	6	7
… make practicing easier?	1	2	3	4	5	6	7
To what degree was this feedback…	**(from 1=low to 7=high)**						
… useful?	1	2	3	4	5	6	7
… easy to learn to operate?	1	2	3	4	5	6	7
… something that did what you wanted it to do?	1	2	3	4	5	6	7
… clear and understandable?	1	2	3	4	5	6	7
… flexible?	1	2	3	4	5	6	7
… easy to become skilled at?	1	2	3	4	5	6	7
… easy to use?	1	2	3	4	5	6	7
… accurate?	1	2	3	4	5	6	7
… something you would use again?	1	2	3	4	5	6	7
… something you would recommend to others?	1	2	3	4	5	6	7
What were the most helpful features?							
What were the biggest weaknesses/limitations?							
How would you improve the feedback?							
Any other feedback on the technology?							

INTRODUCTION BY FACILITATOR

Hello, my name is [facilitator name] and I am with [name of college]. Thank you for taking the time to participate in this focus group about music learning technologies. This focus group is part of a larger needs' assessment process that the TELMI consortium is conducting to learn about the usability and effectiveness of adaptive training technologies for supporting violin learning. TELMI is a three-years project funded by the European Commission, which seeks to engage with and enhance processes of learning the violin.

Particularly, today's focus group is centred on the effectiveness of technologies for the analysis of full-body movement and postural features and of an interface for providing feedback to teachers and students. Technologies are produced by the TELMI consortium partners.

It is generally known, indeed, that the way feedback is presented is particularly crucial in learning technology. This issue assumes a central importance in music learning, since students need to focus on several tasks while training, e.g., their own proprioceptive feedback, music score sheet reading, technique controls, and so on, and the way technology provides its own feedback is crucial to avoid overloading people attention.

To sum up, we want to understand how we might customize an adaptive feedback interface to better fit students' needs.

You are a group of people, here in Genoa and London, who may have an interest in using TELMI technologies, considering particularly the interface for posture and upper body movement control that you already tried. We would like to hear from you about how this interface would best meet your needs for enhancing your violin study and your proficiency.

During this focus group, I will ask you questions and facilitate a conversation between you about how the TELMI technologies you tried might be able to help you with achieving your objectives. Please keep in mind that there are no "right" or "wrong" answers to any of the questions I will ask. The purpose is to stimulate conversation and hear the opinions of everyone in the room. I hope you will be comfortable speaking honestly and sharing your ideas with us.

Please note that this session will be recorded (or [name] will be taking notes during the focus group) to ensure we adequately capture your ideas during the conversation. However, the comments from the focus group will remain confidential and your name will not be attached to any comments you make. Do you have any questions before we start?

FOCUS GROUP QUESTIONS

1. Let's do a quick round of introductions. Can each of you tell the group your name, if you are still studying, if you are an active musician or a retired one, a teacher and/or mainly a performer, what prior experience do you have with learning technologies in general, and what with learning technologies specifically addressed to music learning?
2. In what way are you related to the TELMI Project and to the technologies provided by the TELMI Consortium?
3. *[If the focus group is conducted with music teachers:]* Does the project inspire you to try out new methodologies in your violin teaching?
4. What in your opinion is an important competence in music performance from a biomechanical point of view?
5. What are your current goals as a violin student/teacher, when you consider using such a kind of technology?

a. Probe: Are you looking for an interface that corrects you directly? Are you looking for an interface that visualizes important information for you, such as music scores, posture (e.g., a silhouette that shows you correct postures), or something else?

6. Now, we'd like to hear about the classes in which you think this technology may be successfully applied.
   a. In what ways may this technology be helpful to you and/or in the classes you are considering?
   b. In what ways do you feel that technologies already available in the market (e.g., music apps and software for real-time reproduction of repertoire music sheet, ear training software, and so on) fell short in helping you reach your goals?
7. What are the reported advantages and disadvantages of integrating the TELMI technologies into musical teaching, compared with teaching without such technology?
8. Do you think that technologies such as those you experienced may bring an added value in music learning/teaching? i.e., do you think that using these technologies can make any difference in students learning attitudes and learning outcomes with respect to not using them?
9. Now imagine that you are part of a committee of people designing technologies for music students in Conservatoire age group. These are technologies that people like you might take to advance a music curriculum, or to advance in students' career.
   a. What are the factors that you will make sure your committee will consider in designing this interface? What are the things that you are sure would attract people like you to use this software?
   b. Probe: Remember, these factors can belong to many areas: the violin syllabus that the system should consider, the exercise/repertoire length, the teaching style, body position, rotation, qualities of movement to consider, whether the system should consider the all body or just upper or lower body movements' features, whether the system should promote a better integration and link between body postures and quality of sound production, such as pitch intonation, precision of tuning and so on.
10. *[If this issue has not already been addressed as the questions above were answered:]* At this point we'd like to hear about your direct experience with the TELMI interface for full-body monitoring - testing, advising, and so on. In what ways was the interface helpful to you for self-monitoring in playing a TELMI archive's exercise?
    a. Do you think that the system provided you with a clear **visual** representation of the considered posture and movement qualities?
    b. Do you think that the system may integrate different strategies to stimulate prior information and quicker understanding in violin students?
    c. Do you think that the system may integrate different strategies to facilitate attention maintenance on content being learnt?
    d. Do you think that the system should add some elements to reinforce students' motivation?
    e. Do you think that the system should give students opportunities to further practice difficult passages or position that the system highlighted as bad ones?
    f. Do you think that the system provides easy.to-use navigation tools?

g. Do you think that system feedback was provided in the right sequence and timing while you were playing?

11. *[If this issue has not already been addressed as the questions above were answered:]* In what ways do you feel that the interface fell short in helping you reach your goals or understanding your target movements?
12. We would like to know how to make our system more welcoming and accessible to most of the students and want to hear your thoughts on how we could do that.
13. Is there anything else we haven't discussed yet that you think is important for the TELMI consortium to know about as we consider tailoring adaptive music learning technology to most students?

THANK YOU SO MUCH FOR YOUR TIME!

Bibliography

[tha, 2005] (2005). The future of music in therapy and medicine. *Annals of the New York Academy of Sciences*, 1060(1):303–308.

[Aagten-Murphy et al., 2015] Aagten-Murphy, D., Attucci, C., Daniel, N., Klaric, E., Burr, D., and Pellicano, E. (2015). Numerical estimation in children with autism. *Autism Research*, 8(6):668–681.

[Ahissar and Hochstein, 2004] Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, 8(10):457–464.

[Akhtar and Gernsbacher, 2008] Akhtar, N. and Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child development perspectives*, 2(2):59–65.

[Akilli, 2008] Akilli, G. (2008). Design-based research vs. mixed methods: The differences and commonalities. 1-10.

[Alborno et al., 2016] Alborno, P., Cera, A., Piana, S., Mancini, M., Niewiadomski, R., Canepa, C., Volpe, G., and Camurri, A. (2016). Interactive sonification of movement qualities – a case study on fluidity. In *Proceedings of ISon 2016, 5th Interactive Sonification Workshop*.

[Altenmuller et al., 2009] Altenmuller, E., Marco-Pallares, J., Munte, T., and Schneider, S. (2009). Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. *Annals of the New York Academy of Sciences*, 1169(1):395–405.

[Amengual et al., 2013] Amengual, J. L., Rojo, N., de Las Heras, M. V., Marco-Pallarés, J., Grau-Sánchez, J., Schneider, S., Vaquero, L., Juncadella, M., Montero, J., Mohammadi, B., et al. (2013). Sensorimotor plasticity after music-supported therapy in chronic stroke patients revealed by transcranial magnetic stimulation. *PLoS One*, 8(4):e61883.

[Amory and Seagram, 2003] Amory, A. and Seagram, R. (2003). Educational game models: conceptualization and evaluation: the practice of higher education. *South African Journal of Higher Education*, 17(2):206–217.

[Amunts et al., 1997] Amunts, K., Schlaug, G., Jäncke, L., Steinmetz, H., Schleicher, A., Dabringhaus, A., and Zilles, K. (1997). Motor cortex and hand motor skills: structural compliance in the human brain. *Human brain mapping*, 5(3):206–215.

[An and Tillman, 2014] An, S. and Tillman, D. (2014). Elementary teachers' design of arts based teaching: Investigating the possibility of developing mathematics-music integrated curriculum. *Journal of Curriculum Theorizing*, 30(2).

[An and Capraro, 2011] An, S. A. and Capraro, M. M. (2011). *Music-math integrated activities for elementary and middle school students*. Education for All Irvine, CA.

[An et al., 2011] An, S. A., Ma, T., and Capraro, M. M. (2011). Preservice teachers' beliefs and attitude about teaching and learning mathematics through music: An intervention study. *School Science and Mathematics*, 111(5):236–248.

[An and Tillman, 2015] An, S. A. and Tillman, D. A. (2015). Music activities as a meaningful context for teaching elementary students mathematics: A quasi-experiment time series design with random assigned control group. *European Journal of Science and Mathematics Education*, 3(1):45–60.

[Anderson et al., 2002] Anderson, B. J., Eckburg, P. B., and Relucio, K. I. (2002). Alterations in the thickness of motor cortical subregions after motor-skill learning and exercise. *Learning & memory*, 9(1):1–9.

[Anderson and Winawer, 2005] Anderson, B. L. and Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, 434(7029):79.

[Anderson and Green, 2001] Anderson, M. C. and Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410(6826):366.

[Angeloni and Geffen, 2018] Angeloni, C. and Geffen, M. N. (2018). Contextual modulation of sound processing in the auditory cortex. *Current opinion in neurobiology*, 49:8–15.

[Arbib, 2013] Arbib, M. A. (2013). *Language, music, and the brain: a mysterious relationship*, volume 10. mit Press.

[Ardito et al., 2006] Ardito, C., Costabile, M. F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., and Rossano, V. (2006). An approach to usability evaluation of e-learning applications. *Universal access in the information society*, 4(3):270–283.

[Arroyo et al., 2009] Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009). Emotion sensors go to school. In *AIED*, volume 200, pages 17–24. Citeseer.

[Atherton et al., 2018] Atherton, R. P., Chrobak, Q. M., Rauscher, F. H., Karst, A. T., Hanson, M. D., Steinert, S. W., , and Bowe, K. L. (2018). Shared processing of language and music. *Experimental Psychology*, 65:40–48.

[Ayres and Robbins, 2005] Ayres, A. J. and Robbins, J. (2005). *Sensory integration and the child: Understanding hidden sensory challenges*. Western Psychological Services.

[Baader et al., 2005] Baader, A. P., Kazennikov, O., and Wiesendanger, M. (2005). Coordination of bowing and fingering in violin playing. *Cognitive brain research*, 23(2-3):436–443.

[Baddeley, 2003] Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4(10):829.

[Baddeley and Hitch, 1974] Baddeley, A. D. and Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier.

[Bagui, 1998] Bagui, S. (1998). Reasons for increased learning using multimedia. *Journal of educational multimedia and hypermedia*, 7:3–18.

[Bahrick and Lickliter, 2000] Bahrick, L. E. and Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, 36(2):190.

[Baldis, 2001] Baldis, J. J. (2001). Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 166–173. ACM.

[Ballreich and Baumann, 1996] Ballreich, R. and Baumann, W. (1996). *Grundlagen der Biomechanik des Sports*. Stuttgart. (The basics of biomechanics in sports).

[Bandura, 1997] Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

[Bangert and Schlaug, 2006] Bangert, M. and Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience*, 24(6):1832–1834.

[Barrett and Newell, 2015] Barrett, M. M. and Newell, F. N. (2015). Task-specific, age related effects in the cross-modal identification and localisation of objects. *Multisensory research*, 28(1-2):111–151.

[Barry and McArthur, 1994] Barry, N. H. and McArthur, V. (1994). Teaching practice strategies in the music studio: A survey of applied music teachers. *Psychology of Music*, 22(1):44–55.

[Barsalou, 2008] Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645.

[Barth and Paladino, 2011] Barth, H. C. and Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental science*, 14(1):125–135.

[Bauer and Dammers, 2016] Bauer, W. I. and Dammers, R. J. (2016). Technology in music teacher education: A national survey. *Research Perspectives in Music Education*, 18(1):2–15.

[Bauer and Dunn, 2003] Bauer, W. I. and Dunn, R. E. (2003). The electronic portfolio in music teacher education. *Journal of Music Teacher Education*, 13(1):7–20.

[Bavelier and Neville, 2002] Bavelier, D. and Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nature Reviews Neuroscience*, 3(6):443.

[Bellotti et al., 2010] Bellotti, F., Berta, R., and De Gloria, A. (2010). Designing effective serious games: opportunities and challenges for research. *International Journal of Emerging Technologies in Learning (iJET)*, 5(2010).

[Bengtsson et al., 2009] Bengtsson, S. L., Ullen, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., Forssberg, H., and Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *cortex*, 45(1):62–71.

[Benoit et al., 2000] Benoit, C., Martin, J.-C., Pelachaud, C., Schomaker, L., and Suhm, B. (2000). Audio-visual and multimodal speech systems. *Handbook of Standards and Resources for Spoken Language Systems-Supplement*, 500:1–95.

[Berk and Winsler, 1995] Berk, L. E. and Winsler, A. (1995). *Scaffolding Children's Learning: Vygotsky and Early Childhood Education. NAEYC Research into Practice Series. Volume 7.* ERIC.

[Bermudez and Zatorre, 2005] Bermudez, P. and Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*, 1060(1):395–399.

[Bernsen and Dybkjær, 2009] Bernsen, N. O. and Dybkjær, L. (2009). *Multimodal usability*. Springer Science & Business Media.

[Bertelson and Radeau, 1981] Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & psychophysics*, 29(6):578–584.

[Besson et al., 2007] Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative neurology and neuroscience*, 25(3-4):399–410.

[Bigelow, 2003] Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and Psychopathology*, 15(2):259–275.

[Blomberg et al., 1993] Blomberg, J., Giacomi, J., Mosher, A., and Swenton-Wall, P. (1993). Ethnographic field methods and their relation to design. participatory design: Principles and practices. edited by douglas schuler, aki namioka. hillsdale, nj: L.

[Boltz et al., 2009] Boltz, M. G., Ebendorf, B., and Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception: An Interdisciplinary Journal*, 27(1):43–59.

[Bosch et al., 2016] Bosch, N., D'mello, S. K., Ocumpaugh, J., Baker, R. S., and Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6(2):17.

[Brandfonbrener, 2004] Brandfonbrener, A. G. (2004). Healthier music students: Can medicine and music prescribe in concert? *Medical Problems of Performing Artists*, 19(1):1–2.

[Braun and Clarke, 2006] Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

[Bregman, 1990] Bregman, A. (1990). Auditory scene analysis: The perceptual organization of sound. cambridge, ma, us.

[Broughton and Stevens, 2012] Broughton, M. C. and Stevens, C. J. (2012). Analyzing expressive qualities in movement and stillness: effort-shape analyses of solo marimbists' bodily expression. *Music Perception: An Interdisciplinary Journal*, 29(4):339–357.

[Bruner, 1961] Bruner, J. S. (1961). The act of discovery.

[Bruns and Röder, 2010] Bruns, P. and Röder, B. (2010). Tactile capture of auditory localization is modulated by hand posture. *Experimental Psychology*.

[Bruns et al., 2011] Bruns, P., Spence, C., and Röder, B. (2011). Tactile recalibration of auditory spatial representations. *Experimental Brain Research*, 209(3):333–344.

[Burack, 2005] Burack, J. (2005). Uniting mind and music: Shaw's vision continues. *The American Music Teacher*, 55(1):84.

[Burr et al., 2011] Burr, D., Binda, P., Gori, M., Trommershauser, J., and Körding, K. (2011). Combining information from different senses: dynamic adjustment of combination weights, and the development of cross-modal integration in children. *Book of sensory cue integration*, pages 73–95.

[Camurri et al., 2003] Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., and Volpe, G. (2003). Multimodal analysis of expressive gesture in music and dance performances. In *International gesture workshop*, pages 20–39. Springer.

[Camurri et al., 2005] Camurri, A., Volpe, G., De Poli, G., and Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *Ieee Multimedia*, 12(1):43–53.

[Camurri et al., 2016] Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., and Canepa, C. (2016). The dancer in the eye: Towards a multi-layered computational framework of qualities in movement. In *Proceedings of the 3rd International Symposium on Movement and Computing*, MOCO '16, pages 6:1–6:7, New York, NY, USA. ACM.

[Cao et al., 2018] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

[Carbonell, 1970] Carbonell, J. R. (1970). Ai in cai: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4):190–202.

[Carver et al., 1999] Carver, C. A., Howard, R. A., and Lane, W. D. (1999). Enhancing student learning through hypermedia courseware and incorporation of student learning styles. *IEEE transactions on Education*, 42(1):33–38.

[Case-Smith and O'Brien, 2013] Case-Smith, J. and O'Brien, J. C. (2013). *Occupational therapy for children-E-Book*. Elsevier Health Sciences.

[Castellano et al., 2009] Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, ICMI-MLMI '09, pages 119–126, New York, NY, USA. ACM.

[Chaffin et al., 2010] Chaffin, R., Lisboa, T., Logan, T., and Begosh, K. T. (2010). Preparing for memorized cello performance: The role of performance cues. *Psychology of music*, 38(1):3–30.

[Chandler, 1994] Chandler, D. (1994). Semiotics for beginners.

[Chandrasekaran et al., 2009] Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology*, 5(7):e1000436.

[Chandrashekar et al., 2006] Chandrashekar, S., Stockman, T., Fels, D., and Benedyk, R. (2006). Using think aloud protocol with blind users:: a case for inclusive usability evaluation methods. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 251–252. ACM.

[Chen et al., 1999] Chen, H., Wigand, R. T., and Nilan, M. S. (1999). Optimal experience of web activities. *Computers in human behavior*, 15(5):585–608.

[Cheng et al., 2015] Cheng, M.-T., Chen, J.-H., Chu, S.-J., and Chen, S.-Y. (2015). The use of serious games in science education: a review of selected empirical research from 2002 to 2013. *Journal of computers in education*, 2(3):353–375.

[Chobert et al., 2011] Chobert, J., Marie, C., François, C., Schön, D., and Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience*, 23(12):3874–3887.

[Chorianopoulou et al., 2017] Chorianopoulou, A., Tzinis, E., Iosif, E., Papoulidi, A., Papailiou, C., and Potamianos, A. (2017). Engagement detection for children with autism spectrum disorder. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5055–5059. IEEE.

[Chuang and Kuo, 2016] Chuang, T.-Y. and Kuo, M.-S. (2016). A motion-sensing game-based therapy to foster the learning of children with sensory integration dysfunction. *Journal of Educational Technology & Society*, 19(1):4–16.

[Cicchetti, 1994] Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4):284.

[Clark, 2008] Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. OUP USA.

[Clark and Chalmers, 1998] Clark, A. and Chalmers, D. (1998). The extended mind. *analysis*, 58(1):7–19.

[Clark and Paivio, 1991] Clark, J. M. and Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3):149–210.

[Clements et al., 2011] Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., and Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, 42(2):127–166.

[Cochrane et al., 2013] Cochrane, T., Fantini, B., and Scherer, K. R. (2013). *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control*. OUP Oxford.

[Coffield et al., 2004] Coffield, F., Moseley, D., Hall, E., Ecclestone, K., et al. (2004). Learning styles and pedagogy in post-16 learning: A systematic and critical review.

[Cohen et al., 1997] Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997). Quickset: Multimodal interaction for simulation set-up and control. In *Proceedings of the fifth conference on Applied natural language processing*, pages 20–24. Association for Computational Linguistics.

[Cohen and McGee, 2004] Cohen, P. R. and McGee, D. R. (2004). Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1):41–46.

[Cole and Pheng, 1998] Cole, P. G. and Pheng, L. C. (1998). The effects of verbal mediation training on the problem-solving skills of children with partial sight and children without visual impairments. *International Journal of Disability, Development and Education*, 45(4):411–422.

[Colling et al., 2017] Colling, L. J., Noble, H. L., and Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: an eeg study. *Frontiers in neuroscience*, 11:360.

[Colwell and Abrahams, 1991] Colwell, R. and Abrahams, F. (1991). Edwin gordon's contribution: An appraisal. *The Quarterly Journal of Music Teaching and Learning*, 2(1/2):19–36.

[Condie, 2007] Condie, R. (2007). The impact of ict in schools, a landscape review, becta. *http://publications. becta. org. uk/download. cfm? resID= 28221*.

[Connors et al., 2014] Connors, E., Chrastil, E., Sanchez, J., and Merabet, L. B. (2014). Action video game play and transfer of navigation and spatial cognition skills in adolescents who are blind. *Frontiers in human neuroscience*, 8:133.

[Cools et al., 2008] Cools, R., Gibbs, S. E., Miyakawa, A., Jagust, W., and D'Esposito, M. (2008). Working memory capacity predicts dopamine synthesis capacity in the human striatum. *Journal of Neuroscience*, 28(5):1208–1212.

[Costa and Rocha, 2005] Costa, F. A. e. and Rocha, L. M. (2005). Introduction to the special issue: Embodied and situated cognition.

[Council et al., 2001] Council, N. R., Up, A. I., et al. (2001). Helping children learn mathematics. *Mathematics Learning Study Committee. J. Kilpatrick, J. Swafford, and B. Findell, eds. Washington, DC: National Academy Press*.

[Courey et al., 2012] Courey, S. J., Balogh, E., Siker, J. R., and Paik, J. (2012). Academic music: music instruction to engage third-grade students in learning basic fraction concepts. *Educational studies in mathematics*, 81(2):251–278.

[Cox and Stephens, 2006] Cox, H. and Stephens, L. (2006). The effect of music participation on mathematical achievement and overall academic achievement of high school students. *International Journal of Mathematical Education in Science and Technology*, 37(7):757–763.

[Cozzani et al., 2017] Cozzani, G., Pozzi, F., Dagnino, F. M., Katos, A., and Katsouli, E. (2017). Innovative technologies for intangible cultural heritage education and preservation: the case of i-treasures. *Personal and Ubiquitous Computing*, 21(2):253–265.

[Crabtree, 1998] Crabtree, A. (1998). Ethnography in participatory design. In *Proceedings of the 1998 Participatory design Conference*, pages 93–105. Computer Professionals for Social Responsibility Stanford, CA.

[Cramer et al., 2002] Cramer, K. A., Post, T. R., and delMas, R. C. (2002). Initial fraction learning by fourth-and fifth-grade students: A comparison of the effects of using commercial curricula with the effects of using the rational number project curriculum. *Journal for Research in Mathematics Education*, pages 111–144.

[Cronbach, 1951] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334.

[Csıkszentmihályi, 1975] Csıkszentmihályi, M. (1975). Beyond boredom and anxiety.

[Daffner and Searl, 2008] Daffner, K. R. and Searl, M. M. (2008). The dysexecutive syndromes. *Handbook of clinical neurology*, 88:249–267.

[Damasio and Sutherland, 1994] Damasio, A. and Sutherland, S. (1994). Descartes' error: Emotion, reason and the human brain. *Nature*, 372(6503):287–287.

[Dania et al., 2011] Dania, A., Hatziharistos, D., Koutsouba, M., and Tyrovola, V. (2011). The use of technology in movement and dance education: recent practices and future perspectives. *Procedia-Social and Behavioral Sciences*, 15:3355–3361.

[Darin et al., 2018] Darin, T., Andrade, R., and Sánchez, J. (2018). Clue: A usability evaluation checklist for multimodal video game field studies with children who are blind. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

[Darin et al., 2015] Darin, T., Sánchez, J., and Andrade, R. (2015). Dimensions to analyze the design of multimodal video games for the cognition of people who are blind. In *14th Brazilian Symposium on Human Factors in Computing Systems*, pages 2–11.

[Darin et al., 2017] Darin, T. G., Andrade, R., Merabet, L. B., and Sánchez, J. H. (2017). Investigating the mode in multimodal video games: Usability issues for learners who are blind. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2487–2495. ACM.

[Davidson, 2012] Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, 40(5):595–633.

[De Gelder, 2009] De Gelder, B. (2009). Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484.

[De Vrijer et al., 2008] De Vrijer, M., Medendorp, W. P., and Van Gisbergen, J. (2008). Shared computational mechanism for tilt compensation accounts for biased verticality percepts in motion and pattern vision. *Journal of neurophysiology*, 99(2):915–930.

[Decety and Grèzes, 2006] Decety, J. and Grèzes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain research*, 1079(1):4–14.

[Demos et al., 2014] Demos, A. P., Chaffin, R., and Kant, V. (2014). Toward a dynamical theory of body movement in musical performance. *Frontiers in psychology*, 5:477.

[Dempster and Gaughran, 1967] Dempster, W. T. and Gaughran, G. R. (1967). Properties of body segments based on size and weight. *American journal of anatomy*, 120(1):33–54.

[Desmet et al., 2012] Desmet, F., Nijs, L., Demey, M., Lesaffre, M., Martens, J.-P., and Leman, M. (2012). Assessing a clarinet player's performer gestures in relation to locally intended musical targets. *Journal of New Music Research*, 41(1):31–48.

[Devlin, 2013] Devlin, K. (2013). The music of math games.

[Dewey, 1933] Dewey, J. (1933). How we think. a restatement of the relation of reflective thinking to the educative process, boston etc.(dc heath and company) 1933.

[di Ferdinando et al., 2015] di Ferdinando, A., di Fuccio, R., Ponticorvo, M., and Miglino, O. (2015). Block magic: a prototype bridging digital and physical educational materials to support children learning processes. In *Smart Education and Smart e-Learning*, pages 171–180. Springer.

[Doherty-Sneddon, 2003] Doherty-Sneddon, G. (2003). *Children's unspoken language*. Jessica Kingsley Publishers.

[Doucet et al., 2005] Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagné, J.-P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental brain research*, 160(2):194–202.

[Dourish, 2001] Dourish, P. (2001). Seeking a foundation for context-aware computing. *Human–Computer Interaction*, 16(2-4):229–241.

[Druin, 1999] Druin, A. (1999). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 592–599. ACM.

[Dumas et al., 2009] Dumas, B., Lalanne, D., and Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In *Human machine interaction*, pages 3–26. Springer.

[Durkin et al., 2015] Durkin, K., Boyle, J., Hunter, S., and Conti-Ramsden, G. (2015). Video games for children and adolescents with special educational needs. *Zeitschrift für Psychologie*.

[Eccles, 1999] Eccles, J. S. (1999). The development of children ages 6 to 14. *Future of children*, 9(2):30–44.

[Edelson and Johnson, 2003] Edelson, R. J. and Johnson, G. (2003). Music makes math meaningful. *Childhood Education*, 80(2):65–70.

[Elbert et al., 1995] Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., and Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, 270(5234):305–307.

[Erhel and Jamet, 2013] Erhel, S. and Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67:156–167.

[Ernst and Banks, 2002] Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429.

[EUROPEA, 2013] EUROPEA, U. (2013). Survey of schools: ict in education. *Luxembourg: European Comission*.

[Fahle, 2005] Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current opinion in neurobiology*, 15(2):154–160.

[Fahle et al., 2002] Fahle, M., Poggio, T., Poggio, T. A., et al. (2002). *Perceptual learning*. MIT Press.

[Fautley, 2013] Fautley, M. (2013). The potential of audio and video for formative assessment purposes in music education in the lower secondary school in england: Issues arising from a small-scale study of trainee music teachers. *Journal of Music, Technology & Education*, 6(1):29–42.

[Fauvel et al., 2006] Fauvel, J., Flood, R., Wilson, R., and Behrends, E. (2006). Music and mathematics. *The Mathematical Intelligencer*, 28(3):69–71.

[Fernald and Keller, 1921] Fernald, G. M. and Keller, H. (1921). The effect of kinaesthetic factors in the development of word recognition in the case of non-readers. *The Journal of Educational Research*, 4(5):355–377.

[Finke et al., 1992] Finke, R. A., Ward, T. B., and Smith, S. M. (1992). Creative cognition: Theory, research, and applications.

[Finneran and Zhang, 2003] Finneran, C. M. and Zhang, P. (2003). A person–artefact–task (pat) model of flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, 59(4):475–496.

[Finocchietti et al., 2015] Finocchietti, S., Cappagli, G., Porquis, L. B., Baud-Bovy, G., Cocchi, E., and Gori, M. (2015). Evaluation of the audio bracelet for blind interaction for improving mobility and spatial cognition in early blind children-a pilot study. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7998–8001. IEEE.

[Fischer et al., 2011] Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H.-C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychonomic bulletin & review*, 18(1):177–183.

[Fishbein et al., 1988] Fishbein, M., Middlestadt, S. E., Ottati, V., Straus, S., and Ellis, A. (1988). Medical problems among icsom musicians: overview of a national survey. *Med Probl Perform Art*, 3(1):1–8.

[Flash and Hogan, 1985] Flash, T. and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of neuroscience*, 5(7):1688–1703.

[Flaugnacco et al., 2014] Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., Monasta, L., Montico, M., Sila, A., Ronfani, L., et al. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in human neuroscience*, 8:392.

[for Education, 2006] for Education, D. (2006). *Further Education: Raising Skills, Improving Life Chances*, volume 6768. The Stationery Office.

[for Education, 2013] for Education, D. (2013). Science programmes of study: Key stages 1 and 2. *National curriculum in England: Science programmes of study*.

[Forgeard et al., 2008] Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., and Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, 25(4):383–390.

[Gaab and Schlaug, 2003a] Gaab, N. and Schlaug, G. (2003a). The effect of musicianship on pitch memory in performance matched groups. *Neuroreport*, 14(18):2291–2295.

[Gaab and Schlaug, 2003b] Gaab, N. and Schlaug, G. (2003b). Musicians differ from nonmusicians in brain activation despite performance matching. *Annals of the New York Academy of Sciences*, 999(1):385–388.

[Galamian, 1962] Galamian, I. (1962). Vibrato. *Principles of violin playing and teaching*, pages 37–43.

[Gall, 2013] Gall, M. (2013). Trainee teachers' perceptions: Factors that constrain the use of music technology in teaching placements. *Journal of Music, Technology & Education*, 6(1):5–27.

[Gallace and Spence, 2006] Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, 68(7):1191–1203.

[Garbarini and Adenzato, 2004] Garbarini, F. and Adenzato, M. (2004). At the root of embodied cognition: Cognitive science meets neurophysiology. *Brain and cognition*, 56(1):100–106.

[Gaser and Schlaug, 2003] Gaser, C. and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23(27):9240–9245.

[Gaskins et al., 2007] Gaskins, S., Haight, W., and Lancy, D. F. (2007). The cultural construction of play. *Play and development: Evolutionary, sociocultural, and functional perspectives*, pages 179–202.

[Gaunt et al., 2012] Gaunt, H., Creech, A., Long, M., and Hallam, S. (2012). Supporting conservatoire students towards professional integration: One-to-one tuition and the potential of mentoring. *Music Education Research*, 14(1):25–43.

[Gee, 2008] Gee, J. P. (2008). Learning and games. *The ecology of games: Connecting youth, games, and learning*, 3:21–40.

[Geist et al., 2012] Geist, K., Geist, E. A., and Kuznik, K. (2012). The patterns of music. *Young Children*, 2:75.

[Gerson et al., 2015] Gerson, S. A., Schiavio, A., Timmers, R., and Hunnius, S. (2015). Active drumming experience increases infants' sensitivity to audiovisual synchrony during observed drumming actions. *PLoS One*, 10(6):e0130960.

[Gibbs Jr, 2006] Gibbs Jr, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind & Language*, 21(3):434–458.

[Glowinski et al., 2014] Glowinski, D., Baron, N., Grandjean, D., Ott, T., Shirole, K., Torres-Eliard, K., and Rappaz, M.-A. (2014). Analyzing expressive styles and functions of bodily movement in violinist performance. In *Proceedings of the 2014 International Workshop on Movement and Computing*, page 154. ACM.

[Glowinski et al., 2013] Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., and Doherty, C. (2013). The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Frontiers in psychology*, 4:841.

[Gnecco et al., 2014] Gnecco, G., Glowinski, D., Camurri, A., and Sanguineti, M. (2014). On the detection of the level of attention in an orchestra through head movements. *Int. J. Arts Technol*, 7:316–338.

[Godøy and Leman, 2010] Godøy, R. I. and Leman, M. (2010). *Musical gestures: Sound, movement, and meaning*. Routledge.

[Goebl et al., 2014] Goebl, W., Dixon, S., and Schubert, E. (2014). Quantitative methods: Motion analysis, audio analysis, and continuous response techniques. *Expressiveness in music performance: Empirical approaches across styles and cultures*, pages 221–239.

[Goldin-Meadow, 1999] Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences*, 3(11):419–429.

[Goldin-Meadow et al., 2009] Goldin-Meadow, S., Cook, S. W., and Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, 20(3):267–272.

[Gordon, 1965] Gordon, E. (1965). Th e musical aptitude profile: A new and unique musical aptitude test battery.

[Gordon, 2003] Gordon, E. (2003). *A music learning theory for newborn and young children*. Gia Publications.

[Gordon, 2007] Gordon, E. (2007). *Learning sequences in music: A contemporary music learning theory*. Gia Publications.

[Gori et al., 2008] Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, 18(9):694–698.

[Gori et al., 2010] Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, 20(3):223–225.

[Gori et al., 2013] Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2013). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, 137(1):288–293.

[Gori et al., 2012] Gori, M., Tinelli, F., Sandini, G., Cioni, G., and Burr, D. (2012). Impaired visual size-discrimination in children with movement disorders. *Neuropsychologia*, 50(8):1838–1843.

[Goswami, 2011] Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, 15(1):3–10.

[Goswami et al., 2013] Goswami, U., Huss, M., Mead, N., Fosker, T., and Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: significant longitudinal relations with word reading and reading comprehension. *Cortex*, 49(5):1363–1376.

[Goswami et al., 2002] Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., and Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16):10911–10916.

[Gougoux et al., 2004] Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature*, 430(6997):309.

[Gougoux et al., 2005] Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., and Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS biology*, 3(2):e27.

[Graf et al., 2009] Graf, S., Kinshuk, and Liu, T.-C. (2009). Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach. *Journal of Educational Technology & Society*, 12(4):3–14.

[Graf and Kinshuk, 2007] Graf, S. and Kinshuk, K. (2007). Providing adaptive courses in learning management systems with respect to learning styles. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 2576–2583. Association for the Advancement of Computing in Education (AACE).

[Graziano et al., 1999] Graziano, A. B., Peterson, M., and Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal training. *Neurological research*, 21(2):139–152.

[Gredler, 1996] Gredler, M. E. (1996). 17. educational games and simulations: A technology in search of a (research) paradigm. *Technology*, 39:521–540.

[Green et al., 2016] Green, E., Stroud, L., Bloomfield, S., et al. (2016). Griffiths scales of child development.

[Griffin et al., 2015] Griffin, H. J., Aung, M. S. H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. (2015). Perception and automatic recognition of laughter from whole-body motion: Continuous and categorical perspectives. *IEEE Transactions on Affective Computing*, 6(2):165–178.

[Grosshauser and Tröster, 2014] Grosshauser, T. and Tröster, G. (2014). Musical instrument interaction: development of a sensor fingerboard for string instruments. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, pages 177–180. ACM.

[Gruhn and Rauscher, 2002] Gruhn, W. and Rauscher, F. (2002). The neurobiology of music cognition and learning. *The new handbook of research on music teaching and learning*, pages 445–460.

[Habib et al., 2016] Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., and Besson, M. (2016). Music and dyslexia: a new musical training method to improve reading and related disorders. *Frontiers in psychology*, 7:26.

[Hadjakos et al., 2013] Hadjakos, A., Großhauser, T., and Goebl, W. (2013). Motion analysis of music ensembles with the kinect. In *Conference on New Interfaces for Musical Expression*, pages 106–110.

[Hall and O'Donnell, 2011] Hall, C. V. and O'Donnell, J. T. (2011). Calibrating a bowing checker for violin students. *Journal of Music, Technology & Education*, 3(2-3):125–139.

[Hallam, 2001] Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education*, 18(1):27–39.

[Hallgren, 2012] Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23.

[Harkleroad, 2006] Harkleroad, L. (2006). *The math behind the music*, volume 1. Cambridge University Press.

[Harp and Mayer, 1998] Harp, S. F. and Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of educational psychology*, 90(3):414.

[Harrar et al., 2018] Harrar, V., Aubin, S., Chebat, D.-R., Kupers, R., and Ptito, M. (2018). The multisensory blind brain. In *Mobility of Visually Impaired People*, pages 111–136. Springer.

[Hartson et al., 2001] Hartson, H. R., Andre, T. S., and Williges, R. C. (2001). Criteria for evaluating usability evaluation methods. *International journal of human-computer interaction*, 13(4):373–410.

[Henley, 2011] Henley, D. (2011). Music education in england: a review by darren henley for the department for education and the department for culture, media and sport.

[Hennecke et al., 1995] Hennecke, M. E., Prasad, K. V., and Stork, D. G. (1995). Automatic speech recognition system using acoustic and visual signals. In *Conference Record of The Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1214–1218. IEEE.

[Himonides and Purves, 2010] Himonides, E. and Purves, R. (2010). The role of technology. Institute of Education, University of London.

[Ho et al., 2015] Ho, T. K.-L., Lin, H.-s., Chen, C.-K., and Tsai, J.-L. (2015). Development of a computer-based visualised quantitative learning system for playing violin vibrato. *British Journal of Educational Technology*, 46(1):71–81.

[Hoffman and Novak, 1996] Hoffman, D. L. and Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of marketing*, 60(3):50–68.

[Hogan and Sternad, 2007] Hogan, N. and Sternad, D. (2007). On rhythmic and discrete movements: reflections, definitions and implications for motor control. *Experimental Brain Research*, 181(1):13–30.

[Holyoak, 1990] Holyoak, K. J. (1990). Problem solving. *Thinking: An invitation to cognitive science*, 3:117–146.

[Hopkins, 2014] Hopkins, M. T. (2014). Pilot-testing of new software for measuring string players' instrument tuning skills. *Journal of Music, Technology & Education*, 7(1):5–21.

[Horowitz and Prytulak, 1969] Horowitz, L. M. and Prytulak, L. S. (1969). Redintegrative memory. *Psychological Review*, 76(6):519.

[Hostetter and Alibali, 2008] Hostetter, A. B. and Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, 15(3):495–514.

[Hsu et al., 2014] Hsu, J.-L., Zhen, Y.-L., Lin, T.-C., and Chiu, Y.-S. (2014). Personalized music emotion recognition using electroencephalography (eeg). In *2014 IEEE International Symposium on Multimedia*, pages 277–278. IEEE.

[Hudson, 2011] Hudson, N. J. (2011). Musical beauty and information compression: Complex to the ear but simple to the mind? *BMC research notes*, 4(1):9.

[Hui, 2006] Hui, K. (2006). Mozart effect in preschool children? *Early child development and care*, 176(3-4):411–419.

[Huss et al., 2011] Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex*, 47(6):674–689.

[Hwang et al., 2013] Hwang, G.-J., Sung, H.-Y., Hung, C.-M., Huang, I., et al. (2013). A learning style perspective to investigate the necessity of developing adaptive learning systems. *Educational Technology & Society*, 16(2):188–197.

[Hyde et al., 2009] Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., and Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, 29(10):3019–3025.

[Isabelle Peretz, 2003] Isabelle Peretz, R. Z. (2003). *The Cognitive Neuroscience of Music*. Oxford University Press, Oxford.

[Ivanov and Geake, 2003] Ivanov, V. K. and Geake, J. G. (2003). The mozart effect and primary school children. *Psychology of music*, 31(4):405–413.

[Iverson and Goldin-Meadow, 1997] Iverson, J. M. and Goldin-Meadow, S. (1997). What's communication got to do with it? gesture in children blind from birth. *Developmental psychology*, 33(3):453.

[J. Grafsgaard and Lester., 2015] J. Grafsgaard, S. Lee, B. M. K. B. and Lester., J. (2015). Modeling selfefficacy across age groups with automatically tracked facial expression. page 582–585.

[Jacobs, 1969] Jacobs, C. (1969). Investigation of kinesthetics in violin playing. *journal of Research in Music Education*, 17(1):112–114.

[Jaekl et al., 2012] Jaekl, P., Soto-Faraco, S., and Harris, L. R. (2012). Perceived size change induced by audiovisual temporal delays. *Experimental brain research*, 216(3):457–462.

[James, 2010] James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental science*, 13(2):279–288.

[James and Reifinger, 2018] James, L. and Reifinger, J. (2018). The relationship of pitch sight-singing skills with tonal discrimination, language reading skills, and academic ability in children. *Journal of Research in Music Education*, 0(0):0022429418756029.

[Jensenius et al., 2009] Jensenius, A. R., Skogstad, S., Nymoen, K., Tørresen, J., and Høvin, M. (2009). Reduced displays of multidimensional motion capture data sets of musical performance. In *Proceedings of ESCOM*.

[Jetzschke et al., 2016] Jetzschke, S., Ernst, M. O., Moscatelli, A., and Boeddeker, N. (2016). Going round the bend: Persistent personal biases in walked angles. *Neuroscience letters*, 617:72–75.

[Jones, 2006] Jones, G. (2006). Department for education and skills/department of health good practice guidance on the education of children with autistic spectrum disorder. *Child: care, health and development*, 32(5):543–552.

[Jorgensen, 1998] Jorgensen, E. R. (1998). Modeling aspects of type iv music instructional triads. *Bulletin of the Council for Research in Music Education*, pages 43–56.

[Jørgensen, 2004] Jørgensen, H. (2004). Strategies for individual practice. *Musical excellence: Strategies and techniques to enhance performance*, pages 85–103.

[Juntunen and Hyvönen, 2004] Juntunen, M.-L. and Hyvönen, L. (2004). Embodiment in musical knowing: how body movement facilitates learning within dalcroze eurhythmics. *British Journal of Music Education*, 21(2):199–214.

[Juslin, 2000] Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human perception and performance*, 26(6):1797.

[Kafer and Kennell, 1999] Kafer, H. and Kennell, R. (1999). National survey of high school pianists. *The American Music Teacher*, 49(1):34.

[Kalashnikova et al., 2018] Kalashnikova, M., Escudero, P., and Kidd, E. (2018). The development of fast-mapping and novel word retention strategies in monolingual and bilingual infants. *Developmental science*, 21(6):e12674.

[Kapoor et al., 2007] Kapoor, A., Burleson, W., and Picard, R. W. (2007). Automatic prediction of frustration. *International journal of human-computer studies*, 65(8):724–736.

[Karlsson et al., 2009] Karlsson, J., Liljeström, S., and Juslin, P. N. (2009). Teaching musical expression: effects of production and delivery of feedback by teacher vs. computer on rated feedback quality. *Music education research*, 11(2):175–191.

[Kazennikov and Wiesendanger, 2009] Kazennikov, O. and Wiesendanger, M. (2009). Bimanual coordination of bowing and fingering in violinists—effects of position changes and string changes. *Motor control*, 13(3):297–309.

[Keller and Ripoll, 2001] Keller, J. and Ripoll, H. (2001). Reflective-impulsive style and conceptual tempo in a gross motor task. *Perceptual and Motor Skills*, 92(3):739–749.

[Keller, 1999] Keller, J. M. (1999). Using the arcs motivational process in computer-based instruction and distance education. *New directions for teaching and learning*, 1999(78):37–47.

[Kemple et al., 2004] Kemple, K. M., Batey, J. J., and Hartle, L. C. (2004). Music play: Creating centers for musical play and exploration. *YC Young Children*, 59(4):30.

[Kensing and Blomberg, 1998] Kensing, F. and Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, 7(3-4):167–185.

[Keogh et al., 1981] Keogh, J. F., Griffin, N. S., and Spector, R. (1981). Observer perceptions of movement confidence. *Research quarterly for exercise and sport*, 52(4):465–473.

[Ketamo, 2003] Ketamo, H. (2003). An adaptive geometry game for handheld devices. *Journal of Educational Technology & Society*, 6(1):83–95.

[Kiili, 2005] Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, 8(1):13–24.

[Kim et al., 2008] Kim, R. S., Seitz, A. R., and Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, 3(1):e1532.

[Kim, 2011] Kim, M., R. W. M. . T. J. (2011). Children's gestures and the embodied knowledge of geometry. *International Journal of Science and Mathematics Education*, 9(1):207–238.

[Kirke and Miranda, 2013] Kirke, A. and Miranda, E. R. (2013). An overview of computer systems for expressive music performance. In *Guide to computing for expressive music performance*, pages 1–47. Springer.

[Kivinen and Ristela, 2003] Kivinen, O. and Ristela, P. (2003). From constructivism to a pragmatist conception of learning. *Oxford review of education*, 29(3):363–375.

[Koffka, 1935] Koffka, K. (1935). Principles of gestalt psychology, chap. vi.

[Kolb, 2014] Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

[Kontra et al., 2012] Kontra, C., Goldin-Meadow, S., and Beilock, S. L. (2012). Embodied learning across the life span. *Topics in cognitive science*, 4(4):731–739.

[Kraus and Chandrasekaran, 2010] Kraus, N. and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature reviews neuroscience*, 11(8):599.

[Kruse et al., 2013] Kruse, N. B., Harlos, S. C., Callahan, R. M., and Herring, M. L. (2013). Skype music lessons in the academy: Intersections of music education, applied music and technology. *Journal of Music, Technology & Education*, 6(1):43–60.

[Kynigos et al., 2010] Kynigos, C., Smyrnaiou, Z., and Roussou, M. (2010). Exploring rules and underlying concepts while engaged with collaborative full-body games. In *Proceedings of the 9th International Conference on Interaction Design and Children*, pages 222–225. ACM.

[Laban and Lawrence, 1947] Laban, R. and Lawrence, F. C. (1947). *Effort*. Macdonald & Evans.

[Lagerlöf et al., 2014] Lagerlöf, P., Wallerstedt, C., and Pramling, N. (2014). Playing, new music technology and the struggle with achieving intersubjectivity. *Journal of Music, Technology & Education*, 7(2):199–216.

[Lahav and Mioduser, 2001] Lahav, O. and Mioduser, D. (2001). Multisensory virtual environment for supporting blind persons' acquisition of spatial cognitive mapping–a case study. In *EdMedia+ Innovate Learning*, pages 1046–1051. Association for the Advancement of Computing in Education (AACE).

[Lahav and Mioduser, 2008a] Lahav, O. and Mioduser, D. (2008a). Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. *Computers in Human Behavior*, 24(3):1139–1155.

[Lahav and Mioduser, 2008b] Lahav, O. and Mioduser, D. (2008b). Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind. *International Journal of Human-Computer Studies*, 66(1):23–35.

[Lakoff and Johnson, 1980] Lakoff, G. and Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2):195–208.

[Lakoff and Núñez, 2000] Lakoff, G. and Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. *AMC*, 10(12):720–733.

[Lappe et al., 2008] Lappe, C., Herholz, S. C., Trainor, L. J., and Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *Journal of Neuroscience*, 28(39):9632–9639.

[Lee et al., 2003] Lee, D. J., Chen, Y., and Schlaug, G. (2003). Corpus callosum: musician and gender effects. *Neuroreport*, 14(2):205–209.

[Lehmann and Ericsson, 1997] Lehmann, A. C. and Ericsson, K. A. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology: A Journal of Research in Music Cognition*, 16(1-2):40.

[Lehmann and Murray, 2005] Lehmann, S. and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, 24(2):326–334.

[Leman, 2007] Leman, M. (2007). Embodied music cognition and mediation technologies.

[Leman, 2012] Leman, M. (2012). Musical gestures and embodied cognition. In *Journées d'informatique musicale (JIM-2012)*, pages 5–7. Université de Mons.

[Leman and Maes, 2015] Leman, M. and Maes, P.-J. (2015). The role of embodiment in the perception of music. *Empirical Musicology Review*, 9(3-4):236–246.

[Leman et al., 2003] Leman, M., Vermeulen, V., De Voogdt, L., Taelman, J., Moelants, D., and Lesaffre, M. (2003). Correlation of gestural musical audio cues and perceived expressive qualities. In *International Gesture Workshop*, pages 40–54. Springer.

[Lesaffre et al., 2003] Lesaffre, M., Leman, M., Tanghe, K., De Baets, B., De Meyer, H., and Martens, J.-P. (2003). User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology. In *Proceedings of the Stockholm Music Acoustics Conference*, volume 10, page 801. Citeseer.

[Lessard et al., 1998] Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, 395(6699):278.

[Lew and Campbell, 2005] Lew, J. C.-T. and Campbell, P. S. (2005). Children's natural and necessary musical play: Global contexts, local applications. *Music Educators Journal*, 91(5):57–62.

[Lim et al., 2018] Lim, T., Lee, S., and Ke, F. (2018). Integrating music into math in a virtual reality game: Learning fractions. In *Virtual and Augmented Reality: Concepts, Methodologies, Tools, and Applications*, pages 1122–1141. IGI Global.

[Litman et al., 2005] Litman, J., Hutchins, T., and Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition & Emotion*, 19(4):559–582.

[Löchtefeld et al., 2011] Löchtefeld, M., Gehring, S., Jung, R., and Krüger, A. (2011). guitar: supporting guitar learning through mobile projection. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1447–1452. ACM.

[Longmuir, 1998] Longmuir, F. (1998). Considerations for fitness appraisal, programming, and counselling of individuals with sensory impairments. *Canadian Journal of Applied Physiology*, 23(2):166–184.

[Loomis, 2014] Loomis, J. M. (2014). Three theories for reconciling the linearity of egocentric distance perception with distortion of shape on the ground plane. *Psychology & Neuroscience*, 7(3):245.

[Loomis et al., 2001] Loomis, J. M., Klatzky, R. L., Golledge, R. G., et al. (2001). Navigating without vision: basic and applied research. *Optometry and vision science*, 78(5):282–289.

[Loyau and Aubergé, 2006] Loyau, F. and Aubergé, V. (2006). Expressions outside the talk turn: ethograms of the feeling of thinking. In *5th LREC*, pages 47–50.

[Lu et al., 2008] Lu, H., Zhang, B., Wang, Y., and Leow, W. K. (2008). idvt: An interactive digital violin tutoring system based on audio-visual fusion. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 1005–1006. ACM.

[Lumbreras and Sánchez, 1998] Lumbreras, M. and Sánchez, J. (1998). 3d aural interactive hyperstories for blind children. *International Journal of Virtual Reality 4 (1), 20*, 28.

[Luriia, 1987] Luriia, A. R. (1987). *The mind of a mnemonist: A little book about a vast memory*. Harvard University Press.

[Machover, 1992] Machover, T. (1992). *Hyperinstruments: A Progress Report, 1987-1991*. MIT Media Laboratory.

[Macrides and Angeli, 2018] Macrides, E. and Angeli, C. (2018). Domain-specific aspects of technological pedagogical content knowledge: Music education and the importance of affect. *TechTrends*.

[Maehara and Saito, 2007] Maehara, Y. and Saito, S. (2007). The relationship between processing and storage in working memory span: Not two sides of the same coin. *Journal of memory and language*, 56(2):212–228.

[Maes et al., 2014] Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in psychology*, 4:1008.

[Magill., 2001] Magill., R. A. (2001). *Motor learning concepts and applications*. Boston, MA.

[Mainemelis et al., 2002] Mainemelis, C., Boyatzis, R. E., and Kolb, D. A. (2002). Learning styles and adaptive flexibility: Testing experiential learning theory. *Management learning*, 33(1):5–33.

[Majewska and Sur, 2006] Majewska, A. K. and Sur, M. (2006). Plasticity and specificity of cortical processing networks. *Trends in neurosciences*, 29(6):323–329.

[Mall et al., 2016] Mall, P., Spychiger, M., Vogel, R., and Zerlik, J. (2016). *European Music Portfolio (EMP)-Maths:'Sounding Ways Into Mathematics': Teacher's Handbook*. Universitätsbibliothek Johann Christian Senckenberg.

[Mammarella et al., 2012] Mammarella, I. C., Todeschini, M., Englaro, G., Lucangeli, D., and Cornoldi, C. (2012). *Geometriatest. Prove di valutazione per la scuola primaria e secondaria di primo grado. Con DVD*. Edizioni Erickson.

[Manches and O'Malley, 2016] Manches, A. and O'Malley, C. (2016). The effects of physical manipulatives on children's numerical strategies. *Cognition and Instruction*, 34(1):27–50.

[Mancini et al., 2014] Mancini, M., Varni, G., Niewiadomski, R., Volpe, G., and Camurri, A. (2014). How is your laugh today? In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems*, pages 1855–1860. ACM.

[Marks et al., 1987] Marks, L. E., Hammeal, R. J., Bornstein, M. H., and Smith, L. B. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, pages i–100.

[Marquez-Borbon, 2016] Marquez-Borbon, A. (2016). Perceptual learning and the emergence of performer-instrument interactions with digital music systems. In *Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference*, 2016 Conference Proceedings, pages 422–431, Milan Italy. Paparazzi Press.

[Marwan et al., 2007] Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6):237–329.

[Mayer et al., 2001] Mayer, R. E., Heiser, J., and Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of educational psychology*, 93(1):187.

[Mayer and Moreno, 2002] Mayer, R. E. and Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and instruction*, 12(1):107–119.

[Mayo, 2007] Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, 50(7):30–35.

[Mazzocco and Devlin, 2008] Mazzocco, M. M. and Devlin, K. T. (2008). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental science*, 11(5):681–691.

[McCORMICK and McPHERSON, 2003] McCORMICK, J. and McPHERSON, G. (2003). The role of self-efficacy in a musical performance examination: An exploratory structural equation analysis. *Psychology of Music*, 31(1):37–51.

[McCrindle and Symons, 2000] McCrindle, R. J. and Symons, D. (2000). Audio space invaders. In *3rd International Conference on Disability, Virtual Reality and Associated Technologies, Alghero, Italy*.

[McDonnall et al., 2012] McDonnall, M. C., Cavenaugh, B. S., and Giesen, J. M. (2012). The relationship between parental involvement and mathematics achievement for students with visual impairments. *The Journal of Special Education*, 45(4):204–215.

[McGee and Cohen, 2003] McGee, D. R. and Cohen, P. R. (2003). *Augmenting environments with multimodal interaction*. PhD thesis, Citeseer.

[McGraw and Wong, 1996] McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1):30.

[McPherson and McCormick, 2006] McPherson, G. E. and McCormick, J. (2006). Self-efficacy and music performance. *Psychology of music*, 34(3):322–336.

[McQuiggan et al., 2008] McQuiggan, S. W., Mott, B. W., and Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-adapted Interaction*, 18(1-2):81–123.

[Meltzoff et al., 2007] Meltzoff, A., Brooks, R., Flom, R., Lee, K., and Muir, D. (2007). Gaze following: Its development and significance.

[Meltzoff et al., 2009] Meltzoff, A. N., Kuhl, P. K., Movellan, J., and Sejnowski, T. J. (2009). Foundations for a new science of learning. *science*, 325(5938):284–288.

[Mereu and Kazman, 1997] Mereu, S. W. and Kazman, R. (1997). Audio enhanced 3d interfaces for visually impaired users. *ACM SIGCAPH Computers and the Physically Handicapped*, (57):10–15.

[Merleau-Ponty, 1945] Merleau-Ponty, M. (1945). Phénomenologie de la perception, gallimard, paris (trad. it. fenomenologia della percezione, il saggiatore, milano, 1972).

[Miller, 1956] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

[Miller et al., 1960] Miller, G. A., Galanter, E., and Pribram, K. H. (1960). Plans and the structure of behavior.

[Mitchelmore and White, 2000] Mitchelmore, M. C. and White, P. (2000). Development of angle concepts by progressive abstraction and generalisation. *Educational Studies in Mathematics*, 41(3):209–238.

[Montessori and Holmes, 1912] Montessori, M. and Holmes, H. W. (1912). *The Montessori Method: Scientific Pedagogy as Applied to Child Education in" The Children's Houses"*. Frederick A. Stokes Company.

[Moreno et al., 2008] Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex*, 19(3):712–723.

[Mousavi et al., 1995] Mousavi, S. Y., Low, R., and Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology*, 87(2):319.

[Muchnik et al., 1991] Muchnik, C., Efrati, M., Nemeth, E., Malin, M., and Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. *Scandinavian audiology*, 20(1):19–23.

[Münte et al., 2001] Münte, T. F., Kohlmetz, C., Nager, W., and Altenmüller, E. (2001). Neuroperception: Superior auditory spatial tuning in conductors. *Nature*, 409(6820):580.

[Murray et al., 2005] Murray, M. M., Foxe, J. J., and Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage*, 27(2):473–478.

[Murray et al., 2004] Murray, M. M., Michel, C. M., De Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., and Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, 21(1):125–135.

[Nardini et al., 2008] Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Current biology*, 18(9):689–693.

[Nava et al., 2016] Nava, E., Grassi, M., and Turati, C. (2016). Audio-visual, visuo-tactile and audio-tactile correspondences in preschoolers. *Multisensory research*, 29(1-3):93–111.

[Navarro et al., 1999] Navarro, J. I., Aguilar, M., Alcalde, C., and Howell, R. (1999). Relationship of arithmetic problem solving and reflective—impulsive cognitive styles in third-grade students. *Psychological Reports*, 85(1):179–186.

[Naveda and Leman, 2010] Naveda, L. and Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (tga). *Music Perception: An Interdisciplinary Journal*, 28(1):93–111.

[Neti et al., 2000] Neti, C., Iyengar, G., Potamianos, G., Senior, A., and Maison, B. (2000). Perceptual interfaces for information interaction: Joint processing of audio and visual information for human-computer interaction. In *Sixth International Conference on Spoken Language Processing*.

[Ng and Nesi, 2008] Ng, K. and Nesi, P. (2008). i-maestro: Technology-enhanced learning and teaching for music. In *NIME*, pages 225–228.

[Nichols, 2002] Nichols, C. (2002). The vbow: A virtual violin bow controller for mapping gesture to synthesis with haptic feedback. *Organised Sound*, 7(2):215–220.

[Nicolopoulou, 1993] Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget, vygotsky, and beyond. *Human development*, 36(1):1–23.

[Nielsen-Englyst, 2003] Nielsen-Englyst, L. (2003). Game design for imaginative conceptualisation. In *Proceedings of the international workshop on experimental interactive learning in industrial management*, pages 149–164. Citeseer.

[Niemeyer and Starlinger, 1981] Niemeyer, W. and Starlinger, I. (1981). Do the blind hear better? investigations on auditory processing in congenital or early acquired blindness ii. central functions. *Audiology*, 20(6):510–515.

[Niewiadomski et al., 2019] Niewiadomski, R., Kolykhalova, K., Piana, S., Alborno, P., Volpe, G., and Camurri, A. (2019). Analysis of movement quality in full-body physical activities. *ACM Transaction on Interactive Intelligent Systems*, 9(1):1:1–1:20.

[Nijs and Leman, 2014] Nijs, L. and Leman, M. (2014). Interactive technologies in the instrumental music classroom: A longitudinal study with the music paint machine. *Computers & Education*, 73:40–59.

[Nock et al., 2004] Nock, H. J., Iyengar, G., and Neti, C. (2004). Multimodal processing by finding common cause. *Communications of the ACM*, 47(1):51–56.

[Norman, 1993] Norman, D. (1993). Things that make us smart, defending human attributes in the age of the machine, p. 58.

[Novack et al., 2014] Novack, M. A., Congdon, E. L., Hemani-Lopez, N., and Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, 25(4):903–910.

[Nusir et al., 2013] Nusir, S., Alsmadi, I., Al-Kabi, M., and Sharadgah, F. (2013). Studying the impact of using multimedia interactive programs on children's ability to learn basic math skills. *E-Learning and Digital Media*, 10(3):305–319.

[Nyberg et al., 2000] Nyberg, L., Habib, R., McIntosh, A. R., and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences*, 97(20):11120–11124.

[O'Leary, 2014] O'Leary, Z. (2014). The essential guide to doing your research project, 2nd edithion.

[Olugbade et al., 2018a] Olugbade, T., Berthouze, N., Marquardt, N., and Williams, A. (2018a). Human observer and automatic assessment of movement related self-efficacy in chronic pain: From exercise to functional activity. *IEEE Transactions on Affective Computing*.

[Olugbade et al., 2017] Olugbade, T., Cuturi, L., Cappagli, G., Volta, E., Alborno, P., Newbold, J., Bianchi-Berthouze, N., Baud-Bovy, G., Volpe, G., and Gori, M. (2017). What cognitive and affective states should technology monitor to support learning? In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*, pages 22–25. ACM.

[Olugbade et al., 2018b] Olugbade, T. A., Newbold, J., Johnson, R., Volta, E., Alborno, P., Niewiadomski, R., Dillon, M., Volpe, G., and Bianchi-Berthouze, N. (2018b). Automatic detection of reflective thinking in mathematical problem solving based on unconstrained bodily exploration. *arXiv preprint arXiv:1812.07941*.

[Orton, 1928] Orton, S. T. (1928). Specific reading disability—strephosymbolia. *Journal of the American Medical Association*, 90(14):1095–1099.

[Oulasvirta and Salovaara, 2004] Oulasvirta, A. and Salovaara, A. (2004). A cognitive meta-analysis of design approaches to interruptions in intelligent environments. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1155–1158. ACM.

[Overholt, 2005] Overholt, D. (2005). The overtone violin. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 34–37. National University of Singapore.

[Oviatt, 2003] Oviatt, S. (2003). Advances in robust multimodal interface design. *IEEE computer graphics and applications*, (5):62–68.

[Oviatt, 2006] Oviatt, S. (2006). Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 871–880. ACM.

[Oviatt and Cohen, 2000] Oviatt, S. and Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3):45–53.

[Oviatt and Cohen, 2015] Oviatt, S. and Cohen, P. R. (2015). The paradigm shift to multimodality in contemporary computer interfaces. *Synthesis Lectures On Human-Centered Informatics*, 8(3):1–243.

[Oviatt et al., 2003] Oviatt, S., Coulston, R., Tomko, S., Xiao, B., Lunsford, R., Wesson, M., and Carmichael, L. (2003). Toward a theory of organized multimodal integration patterns during human-computer interaction. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 44–51. ACM.

[Oviatt et al., 2004] Oviatt, S., Darrell, T., and Flickner, M. (2004). Multimodal interfaces that flex adapt, and persist. *Communications of the ACM*, 47(1):30–33.

[Oviatt et al., 2005] Oviatt, S., Lunsford, R., and Coulston, R. (2005). Individual differences in multimodal integration patterns: What are they and why do they exist? In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 241–249. ACM.

[Oviatt et al., 2017] Oviatt, S., Schuller, B., Cohen, P., Sonntag, D., and Potamianos, G. (2017). *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*. Morgan & Claypool.

[Oviatt and Cohen, 2010] Oviatt, S. L. and Cohen, A. O. (2010). Toward high-performance communications interfaces for science problem solving. *Journal of science education and technology*, 19(6):515–531.

[Pajares, 1996] Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of educational research*, 66(4):543–578.

[Pankanti et al., 2000] Pankanti, S., Bolle, R. M., and Jain, A. (2000). Biometrics: The future of identification [guest eeditors' introduction]. *Computer*, 33(2):46–49.

[Papic et al., 2011] Papic, M. M., Mulligan, J. T., and Mitchelmore, M. C. (2011). Assessing the development of preschoolers' mathematical patterning. *Journal for Research in Mathematics Education*, 42(3):237–269.

[Paredes and Rodriguez, 2004] Paredes, P. and Rodriguez, P. (2004). A mixed approach to modelling learning styles in adaptive educational hypermedia. *Advanced Technology for Learning*, 1(4):210–215.

[Paris et al., 1990] Paris, S. G., Winograd, P., et al. (1990). How metacognition can promote academic learning and instruction. *Dimensions of thinking and cognitive instruction*, 1:15–51.

[Parise, 2016] Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1-3):7–28.

[Parise and Spence, 2009] Parise, C. V. and Spence, C. (2009). 'when birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, 4(5):e5664.

[Patel, 2011] Patel, A. D. (2011). Why does musical training benefit the neural encoding of speech? a new hypothesis. *The Journal of the Acoustical Society of America*, 130(4):2398–2398.

[Pecher and Zwaan, 2005] Pecher, D. and Zwaan, R. A. (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

[Pekrun and Linnenbrink-Garcia, 2014] Pekrun, R. and Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In *International handbook of emotions in education*, pages 11–20. Routledge.

[Peretz and Zatorre, 2003] Peretz, I. and Zatorre, R. J. (2003). *The cognitive neuroscience of music*. OUP Oxford.

[Peter Visentin and Wasiak, 2008] Peter Visentin, G. S. and Wasiak, E. B. (2008). Informing music teaching and learning using movement analysis technology. *International Journal Of Music Education*, 26(1):73–87.

[Philbeck et al., 2001] Philbeck, J. W., Klatzky, R. L., Behrmann, M., Loomis, J. M., and Goodridge, J. (2001). Active control of locomotion facilitates nonvisual navigation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1):141.

[Phillips, 1995] Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational researcher*, 24(7):5–12.

[Phillips-Silver and Trainor, 2005] Phillips-Silver, J. and Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science*, 308(5727):1430–1430.

[Piaget, 1972] Piaget, J. (1972). *The Principles of Genetic Epistemology: Translated From the French by Wolfe Mays*. Routledge and K. Paul.

[Piana et al., 2013] Piana, S., Staglianò, A., Camurri, A., and Odone, F. (2013). A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In *IDGEI International Workshop*.

[Pike and Shoemaker, 2013] Pike, P. D. and Shoemaker, K. (2013). The effect of distance learning on acquisition of piano sight-reading skills. *Journal of Music, Technology & Education*, 6(2):147–162.

[Piro and Ortiz, 2009] Piro, J. M. and Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychology of Music*, 37(3):325–347.

[Pissaloux and Velazquez, 2017] Pissaloux, E. and Velazquez, R. (2017). Mobility of visually impaired people. In *Fundamentals and ICT Assistive Technologies*. Springer.

[Pooresmaeili et al., 2013] Pooresmaeili, A., Arrighi, R., Biagi, L., and Morrone, M. C. (2013). Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *Journal of Neuroscience*, 33(40):15999–16008.

[Port, 2003] Port, R. F. (2003). Meter and speech. *Journal of phonetics*, 31(3-4):599–611.

[Power et al., 2016] Power, A. J., Colling, L. J., Mead, N., Barnes, L., and Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, 160:1–10.

[Power et al., 2013] Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in human neuroscience*, 7:777.

[Pozzi et al., 2015] Pozzi, F., Alivizatou, M., Dagnino, F., and Ott, M. (2015). Going beyond preservation: how to support technology-enhanced learning in ich education. *International Journal of Heritage in the Digital Era*, 4(1):21–40.

[Prensky, 2001] Prensky, M. (2001). Why games engage us. *Digital Game-Based Learning*, page 2001.

[Price et al., 2017] Price, S., Duffy, S., and Gori, M. (2017). Developing a pedagogical framework for designing a multisensory serious gaming environment. In *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*, pages 1–9. ACM.

[Purves, 2012] Purves, R. (2012). Technology and the educator. Oxford University Press.

[Quinn, 1994] Quinn, C. N. (1994). Designing educational computer games. In *Proceedings of the IFIP TC3/WG3. 2 Working Conference on the Seign, Implementation and Evaluation of Interactive Multimedia in University Settings: Designing for Change in Teaching and Learning*, pages 45–57. Elsevier Science Inc.

[Radford, 2003] Radford, L. (2003). Gestures, speech, and the sprouting of signs: A semiotic-cultural approach to students' types of generalization. *Mathematical thinking and learning*, 5(1):37–70.

[Radford, 2009] Radford, L. (2009). Why do gestures matter? sensuous cognition and the palpability of mathematical meanings. *Educational Studies in Mathematics*, 70(2):111–126.

[Raisamo et al., 2006] Raisamo, R., Hippula, A., Patomaki, S., Tuominen, E., Pasto, V., and Hasu, M. (2006). Testing usability of multimodal applications with visually impaired children. *IEEE MultiMedia*, 13(3):70–76.

[Rauscher et al., 1997] Rauscher, F., Shaw, G., Levine, L., Wright, E., Dennis, W., and Newcomb, R. (1997). Music training causes long-term enhancement of preschool children's spatial–temporal reasoning. *Neurological research*, 19(1):2–8.

[Rauscher and Hinton, 2006] Rauscher, F. H. and Hinton, S. C. (2006). The mozart effect: Music listening is not music instruction. *Educational Psychologist*, 41(4):233–238.

[Rauscher et al., 1993] Rauscher, F. H., Shaw, G. L., and Ky, C. N. (1993). Music and spatial task performance. *Nature*, 365(6447):611.

[Rauscher et al., 1995] Rauscher, F. H., Shaw, G. L., and Ky, K. N. (1995). Listening to mozart enhances spatial-temporal reasoning: towards a neurophysiological basis. *Neuroscience letters*, 185(1):44–47.

[Rauscher and Zupan, 2000] Rauscher, F. H. and Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early childhood research quarterly*, 15(2):215–228.

[Rehg et al., 2013] Rehg, J., Abowd, G., Rozga, A., Romero, M., Clements, M., Sclaroff, S., Essa, I., Ousley, O., Li, Y., Kim, C., et al. (2013). Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421.

[Reithinger et al., 2003] Reithinger, N., Alexandersson, J., Becker, T., Blocher, A., Engel, R., Löckelt, M., Müller, J., Pfleger, N., Poller, P., Streit, M., et al. (2003). Smartkom: adaptive and flexible multimodal access to multiple applications. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 101–108. ACM.

[Repp, 2005] Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, 12(6):969–992.

[Ritchie and Williamon, 2012] Ritchie, L. and Williamon, A. (2012). Self-efficacy as a predictor of musical performance quality. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4):334.

[Ritterfeld et al., 2009] Ritterfeld, U., Cody, M., and Vorderer, P. (2009). *Serious games: Mechanisms and effects*. Routledge.

[Robson, 2004] Robson, B. (2004). Competition in sport, music and dance. *Medical Problems of Performing Artists*, 19(4):160–165.

[Rocha et al., 2016] Rocha, R., Rego, P. A., Faria, B. M., Reis, L. P., and Moreira, P. M. (2016). A web platform of serious games for cognitive rehabilitation: architecture and usability study. In *New Advances in Information Systems and Technologies*, pages 1085–1095. Springer.

[Rodrigo and Baker, 2009] Rodrigo, M. M. T. and Baker, R. S. (2009). Coarse-grained detection of student frustration in an introductory programming course. In *Proceedings of the fifth international workshop on Computing education research workshop*, pages 75–80. ACM.

[RoÈder et al., 1999] RoÈder, B., Teder-SaÈlejaÈrvi, W., Sterr, A., RoÈsler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, 400(6740):162.

[Rollings and Adams, 2003] Rollings, A. and Adams, E. (2003). *Andrew Rollings and Ernest Adams on game design*. New Riders.

[Roseman, 2013] Roseman, I. J. (2013). Appraisal in the emotion system: Coherence in strategies for coping. *Emotion Review*, 5(2):141–149.

[Rosenberg et al., 2010] Rosenberg, D., Depp, C. A., Vahia, I. V., Reichstadt, J., Palmer, B. W., Kerr, J., Norman, G., and Jeste, D. V. (2010). Exergames for subsyndromal depression in older adults: a pilot study of a novel intervention. *The American Journal of Geriatric Psychiatry*, 18(3):221–226.

[Rosson and Carroll, 2009] Rosson, M. B. and Carroll, J. M. (2009). Scenario-based design. In *Human-computer interaction*, pages 161–180. CRC Press.

[Rozencwajg and Corroyer, 2005] Rozencwajg, P. and Corroyer, D. (2005). Cognitive processes in the reflective-impulsive cognitive style. *The Journal of genetic psychology*, 166(4):451–463.

[Ruggieri and Katsnelson, 1996] Ruggieri, V. and Katsnelson, A. (1996). An analysis of a performance by the violinist d. oistrakh: The hypothetical role of postural tonic-static and entourage movements. *Perceptual and motor skills*, 82(1):291–300.

[Salcedo, 2010] Salcedo, C. S. (2010). The effects of songs in the foreign language classroom on text recall, delayed text recall and involuntary mental rehearsal. *Journal of College Teaching & Learning (TLC)*, 7(6).

[Sánchez et al., 2004] Sánchez, J., Baloian, N., and Flores, H. (2004). A methodology for developing audio-based interactive environments for learners with visual disabilities. In *EdMedia+ Innovate Learning*, pages 540–545. Association for the Advancement of Computing in Education (AACE).

[Sánchez et al., 2015] Sánchez, J., Darin, T., and Andrade, R. (2015). Multimodal videogames for the cognition of people who are blind: trends and issues. In *International Conference on Universal Access in Human-Computer Interaction*, pages 535–546. Springer.

[Sánchez and Elías, 2009] Sánchez, J. and Elías, M. (2009). Science learning in blind children through audio-based games. In *Engineering the User Interface*, pages 1–16. Springer.

[Sánchez and Lumbreras, 2000] Sánchez, J. and Lumbreras, M. (2000). Usability and cognitive impact of the interaction with 3d virtual interactive acoustic environments by blind children. In *Proceedings of the 3rd International Conference on Disability, Virtual Reality and Associated Technologies*, pages 67–73.

[Sánchez et al., 2001] Sánchez, J., Lumbreras, M., and Cernuzzi, L. (2001). Interactive virtual acoustic environments for blind children: computing, usability, and cognition. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, pages 65–66. ACM.

[Sánchez and Olivares, 2011] Sánchez, J. and Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education*, 57(3):1943–1952.

[Sanders, 2002] Sanders, E. B.-N. (2002). From user-centered to participatory design approaches. In *Design and the social sciences*, pages 18–25. CRC Press.

[Sanghvi et al., 2011] Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 305–312. ACM.

[Särkämö et al., 2008] Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., Autti, T., Silvennoinen, H. M., Erkkilä, J., Laine, M., et al. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, 131(3):866–876.

[Schellenberg, 2004] Schellenberg, E. G. (2004). Music lessons enhance iq. *Psychological science*, 15(8):511–514.

[Schlaug, 2001] Schlaug, G. (2001). The brain of musicians: a model for functional and structural adaptation. *Annals of the New York Academy of Sciences*, 930(1):281–299.

[Schlaug et al., 1995] Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., and Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8):1047–1055.

[Schlaug et al., 2005] Schlaug, G., Norton, A., Overy, K., Winner, E., et al. (2005). Effects of music training on the child's brain and cognitive development. *ANNALS-NEW YORK ACADEMY OF SCIENCES*, 1060:219.

[Schneider et al., 2002] Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature neuroscience*, 5(7):688.

[Schön et al., 2007] Schön, D., Akiva-Kabiri, L., and Vecchi, T. (2007). *Psicologia della musica*, volume 265. Carocci.

[Schön et al., 2008] Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, 106(2):975–983.

[Schunk, 1981] Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of educational psychology*, 73(1):93.

[Seitz et al., 2006] Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, 16(14):1422–1427.

[Selker, 2004] Selker, T. (2004). Visual attentive interfaces. *BT Technology Journal*, 22(4):146–150.

[Semin and Smith, 2008] Semin, G. R. and Smith, E. R. (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. Cambridge University Press.

[Shams and Seitz, 2008] Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11):411–417.

[Shimojo and Shams, 2001] Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current opinion in neurobiology*, 11(4):505–509.

[Shin et al., 2012] Shin, N., Sutherland, L. M., Norris, C. A., and Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British journal of educational technology*, 43(4):540–560.

[Siegler et al., 2011] Siegler, R. S., Thompson, C. A., and Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive psychology*, 62(4):273–296.

[Sievers et al., 2013] Sievers, B., Polansky, L., Casey, M., and Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences*, 110(1):70–75.

[Silveira and Gavin, 2016] Silveira, J. M. and Gavin, R. (2016). The effect of audio recording and playback on self-assessment among middle school instrumental music students. *Psychology of Music*, 44(4):880–892.

[Skadberg and Kimmel, 2004] Skadberg, Y. X. and Kimmel, J. R. (2004). Visitors' flow experience while browsing a web site: its measurement, contributing factors and consequences. *Computers in human behavior*, 20(3):403–422.

[Sleeuwenhoek et al., 1995] Sleeuwenhoek, H. C., Boter, R. D., and Vermeer, A. (1995). Perceptual-motor performance and the social development of visually impaired children. *Journal of Visual Impairment & Blindness*.

[Sleigh and Lickliter, 1997] Sleigh, M. J. and Lickliter, R. (1997). Augmented prenatal auditory stimulation alters postnatal perception, arousal, and survival in bobwhite quail chicks. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 30(3):201–212.

[Sloboda and Davidson, 1996] Sloboda, J. A. and Davidson, J. (1996). The young performing musician. *Musical beginnings: Origins and development of musical competence*, pages 171–190.

[Sluming et al., 2002] Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., and Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in broca's area in male symphony orchestra musicians. *Neuroimage*, 17(3):1613–1622.

[Sluming et al., 2007] Sluming, V., Brooks, J., Howard, M., Downes, J. J., and Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *Journal of Neuroscience*, 27(14):3799–3806.

[Smith and Sera, 1992] Smith, L. B. and Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive psychology*, 24(1):99–142.

[Snowling, 1980] Snowling, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology*, 29(2):294–305.

[Sommerville and Woodward, 2010] Sommerville, J. A. and Woodward, A. L. (2010). The link between action production and action processing in infancy. *Naturalizing intention in action*, pages 67–89.

[Southgate and Roscigno, 2009] Southgate, D. E. and Roscigno, V. J. (2009). The impact of music on childhood and adolescent achievement. *Social science quarterly*, 90(1):4–21.

[Spelke, 2008] Spelke, E. (2008). Learning, arts and the brain: The dana consortium report on arts and cognition.

[Spence, 2011] Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4):971–995.

[Spencer et al., 1992] Spencer, C., Morsley, K., Ungar, S., Pike, E., and Blades, M. (1992). Developing the blind child's cognition of the environment: the role of direct and map-given experience. *Geoforum*, 23(2):191–197.

[Spinuzzi, 2005] Spinuzzi, C. (2005). The methodology of participatory design. *Technical communication*, 52(2):163–174.

[Starker and Bekefi, 1965] Starker, J. and Bekefi, G. (1965). *An organized method of string playing: violoncello exercises for the left hand*. Peer International Corp.

[Stein, 2012] Stein, B. E. (2012). *The new handbook of multisensory processing*. Mit Press.

[Strauss and Lehtinen, 1947] Strauss, A. A. and Lehtinen, L. E. (1947). Psychopathology and education of the brain-injured child.

[Swanwick and Runfola, 2002] Swanwick, K. and Runfola, M. (2002). Developmental characteristics of music learners. Oxford University Press.

[Sweller, 1988] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285.

[Sweller, 1994] Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312.

[Tabry et al., 2013] Tabry, V., Zatorre, R. J., and Voss, P. (2013). The influence of vision on sound localization abilities in both the horizontal and vertical planes. *Frontiers in psychology*, 4:932.

[Tallal and Gaab, 2006] Tallal, P. and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in neurosciences*, 29(7):382–390.

[Tânia et al., 2005] Tânia, L., Williamon, A., Zicari, M., and Eiholzer, H. (2005). Mastery through imitation: A preliminary study. *Musicae Scientiae*, 9(1):75–110.

[Thibeault, 2018] Thibeault, M. D. (2018). Learning with sound recordings: A history of suzuki's mediated pedagogy. *Journal of Research in Music Education*, page 0022429418756879.

[Thompson, 2018] Thompson, M. B. (2018). What does music learning look like? In *Distributed Systems*, pages 19–34. SpringerBriefs in Education, Springer, Cham.

[Thorgersen and Zandén, 2014] Thorgersen, K. and Zandén, O. (2014). The internet as teacher. *Journal of Music, Technology & Education*, 7(2):233–244.

[Tierney and Kraus, 2013] Tierney, A. T. and Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and language*, 124(3):225–231.

[Tomlinson and McTighe, 2006] Tomlinson, C. A. and McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. ASCD.

[Tonelli et al., 2015] Tonelli, A., Brayda, L., and Gori, M. (2015). Task-dependent calibration of auditory spatial perception through environmental visual observation. *Frontiers in systems neuroscience*, 9:84.

[Tonelli et al., 2016] Tonelli, A., Gori, M., and Brayda, L. (2016). The influence of tactile cognitive maps on auditory space perception in sighted persons. *Frontiers in psychology*, 7:1683.

[Torrente et al., 2009] Torrente, J., del Blanco, Á., Moreno-Ger, P., Martínez-Ortiz, I., and Fernández-Manjón, B. (2009). Implementing accessibility in educational videogames with. In *Proceedings of the first ACM international workshop on Multimedia technologies for distance learning*, pages 57–66. ACM.

[Treichler, 1967] Treichler, D. (1967). Are you missing the boat in training aid? film and av communication.

[Trewin et al., 2008] Trewin, S., Hanson, V. L., Laff, M. R., and Cavender, A. (2008). Powerup: an accessible virtual world. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 177–184. ACM.

[Tsai et al., 2015] Tsai, C.-H., Kuo, Y.-H., Chu, K.-C., and Yen, J.-C. (2015). Development and evaluation of game-based learning system using the microsoft kinect sensor. *International Journal of Distributed Sensor Networks*, 11(7):498560.

[Turk and Kölsch, 2004] Turk, M. and Kölsch, M. (2004). Perceptual interfaces. *Emerging Topics in Computer Vision, Prentice Hall*.

[Turk and Robertson, 2000] Turk, M. and Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM*, 43(3):32–32.

[Turner-Stokes and Reid, 1999] Turner-Stokes, L. and Reid, K. (1999). Three-dimensional motion analysis of upper limb movement in the bowing arm of string-playing musicians. *Clinical Biomechanics*, 14(6):426–433.

[Vallotton and Ayoub, 2011] Vallotton, C. and Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, 26(2):169–181.

[Van Der Linden et al., 2009] Van Der Linden, J., Schoonderwaldt, E., and Bird, J. (2009). Towards a real-time system for teaching novices correct violin bowing technique. In *2009 IEEE International Workshop on Haptic Audio visual Environments and Games*, pages 81–86. IEEE.

[Van Dyck et al., 2014] Van Dyck, E., Vansteenkiste, P., Lenoir, M., Lesaffre, M., and Leman, M. (2014). Recognizing induced emotions of happiness and sadness from dance.

[VanDerHeyden et al., 2011] VanDerHeyden, A. M., Broussard, C., Snyder, P., George, J., Lafleur, S. M., and Williams, C. (2011). Measurement of kindergartners' understanding of early mathematical concepts. *School Psychology Review*, 40(2).

[Varni et al., 2010] Varni, G., Volpe, G., and Camurri, A. (2010). A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Transactions on Multimedia*, 12(6):576–590.

[Vaughn, 2000] Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of aesthetic education*, 34(3/4):149–166.

[Vercillo et al., 2017] Vercillo, T., Tonelli, A., and Gori, M. (2017). Intercepting a sound without vision. *PloS one*, 12(5):e0177407.

[Viladot et al., 2018] Viladot, L., Hilton, C., Casals, A., Saunders, J., Carrillo, C., Henley, J., González-Martín, C., Prat, M., and Welch, G. (2018). The integration of music and mathematics education in catalonia and england: perspectives on theory and practice. *Music Education Research*, 20(1):71–82.

[Vines et al., 2005] Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Ioana, M., and Levitin, D. J. (2005). Dimensions of emotion in expressive musical performance. *Ann. NY Acad. Sci*, 1060:462–466.

[Visentin et al., 2015] Visentin, P., Li, S., Tardif, G., and Shan, G. (2015). Unraveling mysteries of personal performance style; biomechanics of left-hand position changes (shifting) in violin performance. *PeerJ*, 3:e1299.

[Visi et al., 2015] Visi, F., Coorevits, E., Schramm, R., and Miranda, E. (2015). Instrumental movements of neophytes: analysis of movement periodicities, commonalities and individualities in mimed violin performance. In *11th International Symposium on CMMR*, pages 518–533. The Laboratory of Mechanics and Acoustics.

[Visi et al., 2014] Visi, F., Schramm, R., and Miranda, E. (2014). Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. In *Proceedings of the 2014 International Workshop on Movement and Computing*, page 100. ACM.

[Volpe et al., 2016] Volpe, G., Alborno, P., Camurri, A., Coletta, P., Ghisio, S., Mancini, M., Niewiadomski, R., and Piana, S. (2016). Designing multimodal interactive systems using eyesweb xmi. In *SERVE@ AVI*, pages 49–56.

[Volpe et al., 2017] Volpe, G., Kolykhalova, K., Volta, E., Ghisio, S., Waddell, G., Alborno, P., Piana, S., Canepa, C., and Ramirez-Melendez, R. (2017). A multimodal corpus for technology-enhanced learning of violin playing. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, page 25. ACM.

[Volta et al., 2018a] Volta, E., Alborno, P., Gori, M., and Volpe, G. (2018a). Designing a multisensory social serious-game for primary school mathematics learning. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9. IEEE.

[Volta et al., 2018b] Volta, E., Mancini, M., Varni, G., and Volpe, G. (2018b). Automatically measuring biomechanical skills of violin performance: an exploratory study. In *Proceedings of the 5th International Conference on Movement and Computing*, page 16. ACM.

[Von Kriegstein and Giraud, 2006] Von Kriegstein, K. and Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS biology*, 4(10):e326.

[von Laban, 1975] von Laban, R. (1975). *Modern educational dance*. Princeton Book Co Pub.

[Voss et al., 2004] Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J.-P., and Lepore, F. (2004). Early-and late-onset blind individuals show supra-normal auditory abilities in far-space. *Current Biology*, 14(19):1734–1738.

[Vygotsky, 1978a] Vygotsky, L. (1978a). Interaction between learning and development. *Readings on the development of children*, 23(3):34–41.

[Vygotsky, 1978b] Vygotsky, L. S. (1978b). Mind in society: The development of higher mental processes (e. rice, ed. & trans.).

[Waldron, 2013] Waldron, J. (2013). User-generated content, youtube and participatory culture on the web: Music learning and teaching in two contrasting online communities. *Music Education Research*, 15(3):257–274.

[Wang and Hannafin, 2005] Wang, F. and Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4):5–23.

[Webster et al., 1993] Webster, J., Trevino, L. K., and Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in human behavior*, 9(4):411–426.

[Webster, 2012] Webster, P. R. (2012). Key research in music technology and music teaching and learning. *Journal of Music, Technology & Education*, 4(2-3):115–130.

[Wechsler, 2003] Wechsler, D. (2003). Wechsler intelligence scale for children–fourth edition (wisc-iv). *San Antonio, TX: The Psychological Corporation.*

[Whitebread et al., 2011] Whitebread, D., Bingham, S., Vygotsky, L., Martin-Korpi, B., OECD, Martin-Korpi, B., for the Early Years Foundation Stage, S. F., for the Early Years Foundation Stage, S. F., Tickell, C., Wood, D., et al. (2011). Developmental psychology and early childhood education. *Nursery World*, 17(9):5–100.

[Whitehead, 2002] Whitehead, B. J. (2002). The effect of music-intensive intervention on mathematics scores of middle and high school students.

[Williamon and Valentine, 2002] Williamon, A. and Valentine, E. (2002). The role of retrieval structures in memorizing music. *Cognitive psychology*, 44(1):1–32.

[Willingham et al., 2015] Willingham, D. T., Hughes, E. M., and Dobolyi, D. G. (2015). The scientific status of learning styles theories. *Teaching of Psychology*, 42(3):266–271.

[Wilson, 2002] Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636.

[Winberg and Hellström, 2000] Winberg, F. and Hellström, S.-O. (2000). The quest for auditory direct manipulation: the sonified towers of hanoi. In *Proc. of the 3rd Int. Conf. on Disability, Virtual Reality and Associated Technologies*.

[Winograd et al., 1986] Winograd, T., Flores, F., and Flores, F. F. (1986). *Understanding computers and cognition: A new foundation for design*. Intellect Books.

[Winold and Thelen, 1994] Winold, H. and Thelen, E. (1994). Coordination and control in the bow arm movements of highly skilled cellists. *Ecological Psychology*, 6(1):1–31.

[Wöllner et al., 2012] Wöllner, C., Deconinck, F. J., Parkinson, J., and Hove, M. J. (2012). The perception of prototypical motion: Synchronization is enhanced with quantitatively morphed gestures of musical conductors. *American Psychological Association*, 38(6):1390–1403.

[Wolpert et al., 2011] Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739.

[Wong et al., 2007] Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4):420.

[Woo, 2014] Woo, J.-C. (2014). Digital game-based learning supports student motivation, cognitive success, and performance outcomes. *Journal of Educational Technology & Society*, 17(3):291–307.

[Wood and Terrell, 1998] Wood, C. and Terrell, C. (1998). Poor readers' ability to detect speech rhythm and perceive rapid speech. *British Journal of Developmental Psychology*, 16(3):397–413.

[Yildirim, 2017] Yildirim, E. Y. (2017). The positive effect of digital game based programming on students. In *International Conference on Applied Human Factors and Ergonomics*, pages 300–308. Springer.

[Yin et al., 2004] Yin, J., Dhanik, A., Hsu, D., and Wang, Y. (2004). The creation of a music-driven digital violinist. In *ACM Multimedia*, volume 4.

[Yin et al., 2005] Yin, J., Wang, Y., and Hsu, D. (2005). Digital violin tutor: an integrated system for beginning violin learners. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 976–985. ACM.

[Young et al., 2012] Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., and Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of educational research*, 82(1):61–89.

[Zambone, 1989] Zambone, A. M. (1989). Serving the young child with visual impairments: An overview of disability impact and intervention needs. *Infants & Young Children*, 2(2):11–23.

[Zatorre et al., 2007] Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, 8(7):547.

[Zentner and Eerola, 2010] Zentner, M. and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107(13):5768–5773.

[Zhai et al., 1999] Zhai, S., Morimoto, C., and Ihde, S. (1999). Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 246–253. ACM.

[Zimmerman, 1989] Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology*, 81(3):329.

[Zimmerman, 1990] Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1):3–17.